

Novel polynucleotides and polypeptides in pathogenic mycobacteria and their use as diagnostics, vaccines and targets for chemotherapy.

This invention relates to the novel polynucleotide sequence we have designated "GS" which we have identified in pathogenic mycobacteria. GS is a pathogenicity island within 8kb of DNA comprising a core region of 5.75kb and an adjacent transmissible element within 2.25kb. GS is contained within *Mycobacterium paratuberculosis*, *Mycobacterium avium* subsp. *silvaticum* and some pathogenic isolates of *M. avium*. Functional portions of the core region of GS are also represented by regions with a high degree of homology that we have identified in cosmids containing genomic DNA from *Mycobacterium tuberculosis*.

15 Background to the invention

Mycobacterium tuberculosis (Mtb) is a major cause of global diseases of humans as well as animals. Although conventional methods of diagnosis including microscopy, culture and skin testing exist for the recognition of these diseases, improved methods particularly new immunodiagnostics and DNA-based detection systems are needed. Drugs used to treat tuberculosis are increasingly encountering the problem of resistant organisms. New drugs targeted at specific pathogenicity determinants as well as new vaccines for the prevention and treatment of tuberculosis are required. The importance of Mtb as a global pathogen is reflected in the commitment being made to sequencing the entire genome of this organism. This has generated a large amount of DNA sequence data of genomic DNA within cosmid and other libraries. Although the DNA sequence is known in the art, the functions of the vast majority of these sequences, the proteins they encode, the biological significance of these proteins, and the overall relevance and use of these genes and their products as diagnostics, vaccines and targets for chemotherapy for tuberculous disease, remains entirely unknown.

35 *Mycobacterium avium* subsp. *silvaticum* (Mavs) is a pathogenic mycobacterium causing diseases of animals and birds, but it can

also affect humans. *Mycobacterium paratuberculosis* (*Mptb*) causes chronic inflammation of the intestine in many species of animals including primates and can also cause Crohn's disease in humans. *Mptb* is associated with other chronic inflammatory diseases of humans such as sarcoidosis. Subclinical *Mptb* infection is widespread in domestic livestock and is present in milk from infected animals. The organism is more resistant to pasteurisation than *Mtb* and can be conveyed to humans in retail milk supplies. *Mptb* is also present in water supplies, particularly those contaminated with run-off from heavily grazed pastures. *Mptb* and *Mavs* contain the insertion elements IS900 and IS902 respectively, and these are linked to pathogenicity in these organisms. IS900 and IS902 provide convenient highly specific multi-copy DNA targets for the sensitive detection of these organisms using DNA-based methods and for the diagnosis of infections in animals and humans. Much improvement is however required in the immunodiagnosis of *Mptb* and *Mavs* infections in animals and humans. *Mptb* and *Mavs* are in general, resistant in vivo to standard anti-tuberculous drugs. Although substantial clinical improvements in infections caused by *Mptb*, such as Crohn's disease, may result from treatment of patients with combinations of existing drugs such as Rifabutin, Clarithromycin or Azithromycin, additional effective drug treatments are required. Furthermore, there is an urgent need for effective vaccines for the prevention and treatment of *Mptb* and *Mavs* infections in animals and humans based upon the recognition of specific pathogenicity determinants.

Pathogenicity islands are, in general, 7-9kb regions of DNA comprising a core domain with multiple ORFs and an adjacent transmissible element. The transmissible element also encodes proteins which may be linked to pathogenicity, such as by providing receptors for cellular recognition. Pathogenicity islands are envisaged as mobile packages of DNA which, when they enter an organism, assist in bringing about its conversion from a non-disease-causing to a disease-causing strain.

Description of the Drawings

Figure 1(a) and (b) shows a linear map of the pathogenicity island GS in *Mavs* (Fig 1a) and in *Mptb* (Fig 1b). The main open reading frames are illustrated as ORFs A to H. ORFs A to F are found within the core region of GS. ORFs G and H are encoded by the adjacent transmissible element portion of GS.

Disclosure of the invention

Using a DNA-based differential analysis technology we have discovered and characterised a novel polynucleotide in *Mptb* (isolates 0022 from a Guernsey cow and 0021 from a red deer). This polynucleotide comprises the gene region we have designated GS. GS is found in *Mptb* using the identifier DNA sequences Seq.ID.No 1 and 2 where the Seq.ID No2 is the complementary sequence of Seq.ID No 1. GS is also identified in *Mavs*. The complete DNA sequence incorporating the positive strand of GS from an isolate of *Mavs* comprising 7995 nucleotides, including the core region of GS and adjacent transmissible element, is given in Seq.ID No.3. DNA sequence comprising 4435 bp of the positive strand of GS obtained from an isolate of *Mptb* including the core region of GS (nucleotides 1614 to 6047 of GS in *Mavs*) is given in Seq.ID No 4. The DNA sequence of GS from *Mptb* is highly (99.4%) homologous to GS in *Mavs*. The remaining portion of the DNA sequence of GS in *Mptb*, is readily obtainable by a person skilled in the art using standard laboratory procedures. The entire functional DNA sequence including core region and transmissible element of GS in *Mptb* and *Mavs* as described above, comprise the polynucleotide sequences of the invention.

There are 8 open reading frames (ORFs) in GS. Six of these designated GSA, GSB, GSC, GSD, GSE and GSF are encoded by the core DNA region of GS which, characteristically for a pathogenicity island, has a different GC content than the rest of the microbial genome. Two ORFs designated GSG and GSH are encoded by the transmissible element of GS whose GC content resembles that of the rest of the mycobacterial genome. The ORF GSH comprises two sub-ORFs H₁ H₂ on the complementary DNA strand linked by a programmed frameshifting site so that a single polypeptide is translated from the ORF GSH. The nucleotide

sequences of the 8 ORFs in GS and their translations are shown in Seq. ID No 5 to Seq. ID No 29 as follows:

- 5 ORF A: Seq. ID No 5 Nucleotides 50 to 427 of GS from Mavs
Seq. ID No 6 Amino acid sequence encoded by Seq. ID No 5.
- ORF B: Seq. ID No 7 Nucleotides 772 to 1605 of GS from Mavs
Seq. ID No 8 Amino acid sequence encoded by Seq. ID No 7.
- 10 ORF C: Seq. ID No 9 Nucleotides 1814 to 2845 of GS from Mavs
Seq. ID No 10 Amino acid sequence encoded by Seq. ID No 9.
Seq. ID No 11 Nucleotides 201 to 1232 of GS from Mptb
Seq. ID No 12 Amino acid sequence encoded by Seq. ID No 11
- 15 ORF D: Seq. ID No 13 Nucleotides 2785 to 3804 of GS from Mavs
Seq. ID No 14 Amino acid sequence encoded by Seq. ID No 13.
Seq. ID No 15 Nucleotides 1172 to 2191 of GS from Mptb
Seq. ID No 16 Amino acid sequence encoded by Seq. ID No 15.
- 20 ORF E: Seq. ID No 17 Nucleotides 4080 to 4802 of GS from Mavs
Seq. ID No 18 Amino acid sequence encoded by Seq. ID No 17.
Seq. ID No 19 Nucleotides 2467 to 3189 of GS from Mptb
Seq. ID No 20 Amino acid sequence encoded by Seq. ID No 19.
- 25 ORF F: Seq. ID No 21 Nucleotides 4947 to 5747 of GS from Mavs
Seq. ID No 22 Amino acid sequence encoded by Seq. ID No 21.
Seq. ID No 23 Nucleotides 3335 to 4135 of GS from Mptb
Seq. ID No 24 Amino acid sequence encoded by Seq. ID No 23.
- 30

- 5 -

ORF G: Seq. ID No 25 Nucleotides 6176 to 7042 of GS from *Mavs*
Seq. ID No 26 Amino acid sequence encoded by
Seq.ID No 25.

ORF H: Seq.ID No 27 Nucleotides 7953 to 6215 from *Mavs*.

5 ORF H₁: Seq.ID No 28 Amino acid sequence encoded by
nucleotides 7953 to 7006 of Seq.ID No 27

ORF H₂: Seq.ID No 29 Amino acid sequence encoded by
nucleotides 7009 to 6215 of Seq.ID No 27

10 The polynucleotides in *Mtb* with homology to the ORFs B, C, E and
F of GS in *Mptb* and *Mavs*, and the polypeptides they are now known
to encode as a result of our invention, are as follows:

ORF B: Seq.ID No 30 Cosmid MTCY277 nucleotides 35493 to
34705
Seq.ID No 31 Amino acid sequence encoded by Seq.ID
15 No30.

ORF C: Seq.ID No 32 Cosmid MTCY277 nucleotides 31972 to 32994
Seq.ID No 33 Amino acid sequence encoded by Seq.ID
No32.

20 ORF E: Seq.ID No 34 Cosmid MTCY277 nucleotides 34687 to 33956
Seq.ID No 35 Amino acid sequence encoded by Seq.ID
No34.

ORF E: Seq.ID No 36 Cosmid MTO24 nucleotides 15934 to 15203
Seq.ID No 37 Amino acid sequence encoded by Seq.ID
No36.

25 ORF F: Seq.ID No38 Cosmid MTO24 nucleotides 15133 to 14306
Seq.ID No 39 Amino acid sequence encoded by Seq.ID
No38.

The proteins and peptides encoded by the ORFs A to H in *Mptb* and
Mavs and the amino acid sequences from homologous genes we have

discovered in *Mtb* given in Seq.ID Nos 31, 33, 35, 37 and 39, as described above and fragments thereof, comprise the polypeptides of the invention. The polypeptides of the invention are believed to be associated with specific immunoreactivity and with the pathogenicity of the host micro-organisms from which they were obtained.

The present invention thus provides a polynucleotide in substantially isolated form which is capable of selectively hybridising to sequence ID Nos 3 or 4 or a fragment thereof. The polynucleotide fragment may alternatively comprise a sequence selected from the group of Seq.ID.No: 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25 and 27. The invention further provides a polynucleotide in substantially isolated form whose sequence consists essentially of a sequence selected from the group Seq ID Nos. 30, 32, 34, 36 and 38, or a corresponding sequence selectively hybridizable thereto, or a fragment of said sequence or corresponding sequence.

The invention further provides diagnostic probes such as a probe which comprises a fragment of at least 15 nucleotides of a polynucleotide of the invention, or a peptide nucleic acid or similar synthetic sequence specific ligand, optionally carrying a revealing label. The invention also provides a vector carrying a polynucleotide as defined above, particularly an expression vector.

The invention further provides a polypeptide in substantially isolated form which comprises any one of the sequences selected from the group consisting Seq.ID.No: 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 29, 31, 33, 35, 37 and 39, or a polypeptide substantially homologous thereto. The invention additionally provides a polypeptide fragment which comprises a fragment of a polypeptide defined above, said fragment comprising at least 10 amino acids and an epitope. The invention also provides polynucleotides in substantially isolated form which encode polypeptides of the invention, and vectors which comprise such polynucleotides, as well as antibodies capable of binding such polypeptides. In an additional aspect, the invention provides

kits comprising polynucleotides, polypeptides, antibodies or synthetic ligands of the invention and methods of using such kits in diagnosing the presence or absence of mycobacteria in a sample. The invention also provides pharmaceutical compositions comprising polynucleotides of the invention, polypeptides of the invention or antisense probes and the use of such compositions in the treatment or prevention of diseases caused by mycobacteria. The invention also provides polynucleotide prevention and treatment of infections due to GS-containing pathogenic mycobacteria in animals and humans and as a means of enhancing in vivo susceptibility of said mycobacteria to antimicrobial drugs. The invention also provides bacteria or viruses transformed with polynucleotides of the invention for use as vaccines. The invention further provides *Mptb* or *Mavs* in which all or part of the polynucleotides of the invention have been deleted or disabled to provide mutated organisms of lower pathogenicity for use as vaccines in animals and humans. The invention further provides *Mtb* in which all or part of the polynucleotides encoding polypeptides of the invention have been deleted or disabled to provide mutated organisms of lower pathogenicity for use as vaccines in animals and humans.

A further aspect of the invention is our discovery of homologies between the ORFs B, C and E in GS on the one hand, and *Mtb* cosmid MTCY277 on the other (data from Genbank database using the computer programmes BLAST and BLIXEM). The homologous ORFs in MTCY277 are adjacent to one another consistent with the form of another pathogenicity island in *Mtb*. A further aspect of the invention is our discovery of homologies between ORFs E and F in GS, and *Mtb* cosmid MTO24 (also Genbank, as above) with the homologous ORFs close to one another. The use of polynucleotides and polypeptides from *Mtb* (Seq. ID Nos 30, 31, 32, 33, 34, 35, 36, 37, 38 and 39) in substantially isolated form as diagnostics, vaccines and targets for chemotherapy, for the management and prevention of *Mtb* infections in humans and animals, and the processes involved in the preparation and use of these diagnostics, vaccines and new chemotherapeutic agents, comprise further aspects of the invention.

Detailed description of the invention.A. Polynucleotides

Polynucleotides of the invention as defined herein may comprise DNA or RNA. They may also be polynucleotides which include within them synthetic or modified nucleotides or peptide nucleic acids. A number of different types of modification to oligonucleotides are known in the art. These include methylphosphonate and phosphorothioate backbones, addition of acridine or polylysine chains at the 3' and/or 5' ends of the molecule. For the purposes of the present invention, it is to be understood that the polynucleotides described herein may be modified by any method available in the art. Such modifications may be carried out in order to couple the said polynucleotide to a solid phase or to enhance the recognition, the *in vivo* activity, or the lifespan of polynucleotides of the invention.

A number of different types of polynucleotides of the invention are envisaged. In the broadest aspect, polynucleotides and fragments thereof capable of hybridizing to SEQ ID NO:3 or 4 form a first aspect of the invention. This includes the polynucleotide of SEQ ID NO: 3 or 4. Within this class of polynucleotides various sub-classes of polynucleotides are of particular interest.

One sub-class of polynucleotides which is of interest is the class of polynucleotides encoding the open reading frames A, B, C, D, E, F, G and H, including SEQ ID NOs:5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25 and 27. As discussed below, polynucleotides encoding ORF H include the polynucleotide sequences 7953 to 7006 and 7009 to 6215 within SEQ ID NO: 27, as well as modified sequences in which the frame-shift has been modified so that the two sub-reading frames are placed in a single reading frame. This may be desirable where the polypeptide is to be produced in recombinant expression systems.

The invention thus provides a polynucleotide in substantially isolated form which encodes any one of these ORFs or combinations

thereof. Combinations thereof includes combinations of 2, 3, 4, 5 or all of the ORFs. Polynucleotides may be provided which comprise an individual ORF carried in a recombinant vector including the vectors described herein. Thus in one preferred aspect the invention provides a polynucleotide in substantially isolated form capable of selectively hybridizing to the nucleic acid comprising ORFs A to F of the core region of the *Mptb* and *Mavs* pathogenicity islands of the invention. Fragments thereof corresponding to ORFs A to E, B to F, A to D, B to E, A to C, B to D or any two adjacent ORFs are also included in the invention.

Polynucleotides of the invention will be capable of selectively hybridizing to the corresponding portion of the GS region, or to the corresponding ORFs of *Mtb* described herein. The term "selectively hybridizing" indicates that the polynucleotides will hybridize, under conditions of medium to high stringency (for example 0.03 M sodium chloride and 0.03 M sodium citrate at from about 50°C to about 60°C) to the corresponding portion of SEQ ID NO:3 or 4 or the complementary strands thereof but not to genomic DNA from mycobacteria which are usually non-pathogenic including non-pathogenic species of *M.avium*. Such polynucleotides will generally be generally at least 68%, e.g. at least 70%, preferably at least 80 or 90% and more preferably at least 95% homologous to the corresponding DNA of GS. The corresponding portion will be of over a region of at least 20, preferably at least 30, for instance at least 40, 60 or 100 or more contiguous nucleotides.

By "corresponding portion" it is meant a sequence from the GS region of the same or substantially similar size which has been determined, for example by computer alignment, to have the greatest degree of homology to the polynucleotide.

Any combination of the above mentioned degrees of homology and minimum sizes may be used to define polynucleotides of the invention, with the more stringent combinations (i.e. higher homology over longer lengths) being preferred. Thus for example a polynucleotide which is at least 80% homologous over 25, preferably 30 nucleotides forms one aspect of the invention, as

- 10 -

does a polynucleotide which is at least 90% homologous over 40 nucleotides.

A further class of polynucleotides of the invention is the class of polynucleotides encoding polypeptides of the invention, the polypeptides of the invention being defined in section B below. Due to the redundancy of the genetic code as such, polynucleotides may be of a lower degree of homology than required for selective hybridization to the GS region. However, when such polynucleotides encode polypeptides of the invention these polynucleotides form a further aspect. It may for example be desirable where polypeptides of the invention are produced recombinantly to increase the GC content of such polynucleotides. This increase in GC content may result in higher levels of expression via codon usage more appropriate to the host cell in which recombinant expression is taking place.

An additional class of polynucleotides of the invention are those obtainable from cosmids MTCY277 and MT024 (containing *Mtb* genomic sequences), which polynucleotides consist essentially of the fragment of the cosmid containing an open reading frame encoding any one of the homologous ORFs B, C, E or F respectively. Such polynucleotides are referred to below as *Mtb* polynucleotides. However, where reference is made to polynucleotides in general such reference includes *Mtb* polynucleotides unless the context is explicitly to the contrary. In addition, the invention provides polynucleotides which encode the same polypeptide as the abovementioned ORFs of *Mtb* but which, due to the redundancy of the genetic code, have different nucleotide sequences. These form further *Mtb* polynucleotides of the invention. Fragments of *Mtb* polynucleotides suitable for use as probes or primers also form a further aspect of the invention.

The invention further provides polynucleotides in substantially isolated form capable of selectively hybridizing (where selectively hybridizing is as defined above) to the *Mtb* polynucleotides of the invention.

The invention further provides the *Mtb* polynucleotides of the invention linked, at either the 5' and/or 3' end to polynucleotide sequences to which they are not naturally contiguous. Such sequences will typically be sequences found in cloning or expression vectors, such as promoters, 5' untranslated
5 sequence, 3' untranslated sequence or termination sequences. The sequences may also include further coding sequences such as signal sequences used in recombinant production of proteins.

Further polynucleotides of the invention are illustrated in the
10 accompanying examples.

Polynucleotides of the invention may be used to produce a primer, e.g. a PCR primer, a primer for an alternative amplification reaction, a probe e.g. labelled with a revealing label by conventional means using radioactive or non-radioactive labels
15 or a probe linked covalently to a solid phase, or the polynucleotides may be cloned into vectors. Such primers, probes and other fragments will be at least 15, preferably at least 20, for example at least 25, 30 or 40 or more nucleotides in length, and are also encompassed by the term polynucleotides
20 of the invention as used herein.

Primers of the invention which are preferred include primers directed to any part of the ORFs defined herein. The ORFs from other isolates of pathogenic mycobacteria which contain a GS region may be determined and conserved regions within each
25 individual ORF may be identified. Primers directed to such conserved regions form a further preferred aspect of the invention. In addition, the primers and other polynucleotides of the invention may be used to identify, obtain and isolate ORFs capable of selectively hybridizing to the polynucleotides of the
30 invention which are present in pathogenic mycobacteria but which are not part of a pathogenicity island in that particular species of bacteria. Thus in addition to the ORFs B, C, E and F which have been identified in *Mtb*, similar ORFs may be identified in other pathogens and ORFs corresponding to the GS ORFs C, D, E,
35 F and H, may also be identified.

- 12 -

Polynucleotides such as DNA polynucleotides and probes according to the invention may be produced recombinantly, synthetically, or by any means available to those of skill in the art. They may also be cloned by standard techniques.

- 5 In general, primers will be produced by synthetic means, involving a step-wise manufacture of the desired nucleic acid sequence one nucleotide at a time. Techniques for accomplishing this using automated techniques are readily available in the art. Longer polynucleotides will generally be produced using
10 recombinant means, for example using a PCR (polymerase chain reaction) cloning techniques. This will involve making a pair or primers (e.g. of about 15-30 nucleotides) to a region of GS, which it is desired to clone, bringing the primers into contact with genomic DNA from a mycobacterium or a vector carrying the
15 GS sequence, performing a polymerase chain reaction under conditions which bring about amplification of the desired region, isolating the amplified fragment (e.g. by purifying the reaction mixture on an agarose gel) and recovering the amplified DNA. The primers may be designed to contain suitable restriction enzyme
20 recognition sites so that the amplified DNA can be cloned into a suitable cloning vector.

Such techniques may be used to obtain all or part of the GS or ORF sequences described herein, as well as further genomic clones containing full open reading frames. Although in general such
25 techniques are well known in the art, reference may be made in particular to Sambrook J., Fritsch EF., Maniatis T (1989). Molecular cloning: a Laboratory Manual, 2nd edn. Cold Spring Harbor, New York, Cold Spring Harbor Laboratory.

Polynucleotides which are not 100% homologous to the sequences
30 of the present invention but fall within the scope of the invention can be obtained in a number of ways.

Other isolates or strains of pathogenic mycobacteria will be expected to contain allelic variants of the GS sequences described herein, and these may be obtained for example by
35 probing genomic DNA libraries made from such isolates or strains

of bacteria using GS or ORF sequences as probes under conditions of medium to high stringency (for example 0.03M sodium chloride and 0.03M sodium citrate at from about 50°C to about 60°C).

- A particularly preferred group of pathogenic mycobacteria are isolates of *M.paratuberculosis*. Polynucleotides based on GS regions from such bacteria are particularly preferred. Preferred fragments of such regions include fragments encoding individual open reading frames including the preferred groups and combinations of open reading frames discussed above.
- 10 Alternatively, such polynucleotides may be obtained by site directed mutagenesis of the GS or ORF sequences or allelic variants thereof. This may be useful where for example silent codon changes are required to sequences to optimise codon preferences for a particular host cell in which the
- 15 polynucleotide sequences are being expressed. Other sequence changes may be desired in order to introduce restriction enzyme recognition sites, or to alter the property or function of the polypeptides encoded by the polynucleotides of the invention. Such altered property or function will include the addition of
- 20 amino acid sequences of consensus signal peptides known in the art to effect transport and secretion of the modified polypeptide of the invention. Another altered property will include metagenesis of a catalytic residue or generation of fusion proteins with another polypeptide. Such fusion proteins may be
- 25 with an enzyme, with an antibody or with a cytokine or other ligand for a receptor, to target a polypeptide of the invention to a specific cell type in vitro or in vivo.

The invention further provides double stranded polynucleotides comprising a polynucleotide of the invention and its complement.

- 30 Polynucleotides or primers of the invention may carry a revealing label. Suitable labels include radioisotopes such as ³²P or ³⁵S, enzyme labels, other protein labels or smaller labels such as biotin or fluorophores. Such labels may be added to polynucleotides or primers of the invention and may be detected
- 35 using by techniques known per se.

Polynucleotides or primers of the invention or fragments thereof labelled or unlabelled may be used by a person skilled in the art in nucleic acid-based tests for the presence or absence of *Mptb*, *Mavs*, other GS-containing pathogenic mycobacteria, or *Mtb* applied to samples of body fluids, tissues, or excreta from animals and humans, as well as to food and environmental samples such as river or ground water and domestic water supplies.

Human and animal body fluids include sputum, blood, serum, plasma, saliva, milk, urine, csf, semen, faeces and infected discharges. Tissues include intestine, mouth ulcers, skin, lymph nodes, spleen, lung and liver obtained surgically or by a biopsy technique. Animals particularly include commercial livestock such as cattle, sheep, goats, deer, rabbits but wild animals and animals in zoos may also be tested.

Such tests comprise bringing a human or animal body fluid or tissue extract, or an extract of an environmental or food sample, into contact with a probe comprising a polynucleotide or primer of the invention under hybridising conditions and detecting any duplex formed between the probe and nucleic acid in the sample. Such detection may be achieved using techniques such as PCR or by immobilising the probe on a solid support, removing nucleic acid in the sample which is not hybridized to the probe, and then detecting nucleic acid which has hybridized to the probe. Alternatively, the sample nucleic acid may be immobilized on a solid support, and the amount of probe bound to such a support can be detected. Suitable assay methods of this any other formats can be found in for example WO89/03891 and WO90/13667.

Polynucleotides of the invention or fragments thereof labelled or unlabelled may also be used to identify and characterise different strains of *Mptb*, *Mavs*, other GS-containing pathogenic mycobacteria, or *Mtb*, and properties such as drug resistance or susceptibility.

The probes of the invention may conveniently be packaged in the form of a test kit in a suitable container. In such kits the probe may be bound to a solid support where the assay format for

- 15 -

which the kit is designed requires such binding. The kit may also contain suitable reagents for treating the sample to be probed, hybridising the probe to nucleic acid in the sample, control reagents, instructions, and the like.

- 5 The use of polynucleotides of the invention in the diagnosis of inflammatory diseases such as Crohn's disease or sarcoidosis in humans or Johne's disease in animals form a preferred aspect of the invention. The polynucleotides may also be used in the prognosis of these diseases. For example, the response of a
10 human or animal subject in response to antibiotic, vaccination or other therapies may be monitored by utilizing the diagnostic methods of the invention over the course of a period of treatment and following such treatment.

The use of *Mtb* polynucleotides (particularly in the form of
15 probes and primers) of the invention in the above-described methods form a further aspect of the invention, particularly for the detection, diagnosis or prognosis of *Mtb* infections.

B. Polypeptides.

- Polypeptides of the invention include polypeptides in
20 substantially isolated form encoded by GS. This includes the full length polypeptides encoded by the positive and complementary negative strands of GS. Each of the full length polypeptides will contain one of the amino acid sequences set out in Seq ID NOs:6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28 and
25 29. Polypeptides of the invention further include variants of such sequences, including naturally occurring allelic variants and synthetic variants which are substantially homologous to said polypeptides. In this context, substantial homology is regarded as a sequence which has at least 70%, e.g. 80%, 90%, 95% or 98%
30 amino acid homology (identity) over 30 or more, e.g. 40, 50 or 100 amino acids. For example, one group of substantially homologous polypeptides are those which have at least 95% amino acid identity to a polypeptide of any one of Seq ID NOs:6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28 and 29 over their entire length.
35 Even more preferably, this homology is 98%.

Polypeptides of the invention further include the polypeptide sequences of the homologous ORFs of *Mtb*, namely Seq ID Nos. 31, 33, 35, 37 and 39. Unless explicitly specified to the contrary, reference to polypeptides of the invention and their fragments
5 include these *Mtb* polypeptides and fragments, and variants thereof (substantially homologous to said sequences) as defined herein.

Polypeptides of the invention may be obtained by the standard techniques mentioned above. Poly peptides of the invention also
10 include fragments of the above mentioned full length polypeptides and variants thereof, including fragments of the sequences set out in SEQ ID NOS:6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 29, 31, 33, 35, 37 and 39. Such fragments for example of 8, 10, 12, 15 or up to 30 or 40 amino acids may also be obtained
15 synthetically using standard techniques known in the art.

Preferred fragments include those which include an epitope, especially an epitope which is specific to the pathogenicity of the mycobacterial cell from which the polypeptide is derived. Suitable fragments will be at least about 5, e.g. 8, 10, 12, 15
20 or 20 amino acids in size, or larger. Epitopes may be determined either by techniques such as peptide scanning techniques as described by Geysen et al, Mol.Immunol., 23; 709-715 (1986), as well as other techniques known in the art.

The term "an epitope which is specific to the pathogenicity of the mycobacterial cell" means that the epitope is encoded by a
25 portion of the GS region, or by the corresponding ORF sequences of *Mtb* which can be used to distinguish mycobacteria which are pathogenic by from related non-pathogenic mycobacteria including non-pathogenic species of *M.avium*. This may be determined using
30 routine methodology. A candidate epitope from an ORF may be prepared and used to immunise an animal such as a rat or rabbit in order to generate antibodies. The antibodies may then be used to detect the presence of the epitope in pathogenic mycobacteria and to confirm that non-pathogenic mycobacteria do not contain
35 any proteins which react with the epitope. Epitopes may be linear or conformational.

Polypeptides of the invention may be in a substantially isolated form. It will be understood that the polypeptide may be mixed with carriers or diluents which will not interfere with the intended purpose of the polypeptide and still be regarded as substantially isolated. A polypeptide of the invention may also be in a substantially purified form, in which case it will generally comprise the polypeptide in a preparation in which more than 90%, e.g. 95%, 98% or 99% of the polypeptide in the preparation is a polypeptide of the invention.

- 10 Polypeptides of the invention may be modified to confer a desired property or function for example by the addition of Histidine residues to assist their purification or by the addition of a signal sequence to promote their secretion from a cell.

Thus, polypeptides of the invention include fusion proteins which comprise a polypeptide encoding all or part of one or more of an ORF of the invention fused at the N- or C-terminus to a second sequence to provide the desired property or function. Sequences which promote secretion from a cell include, for example the yeast α -factor signal sequence.

- 20 A polypeptide of the invention may be labelled with a revealing label. The revealing label may be any suitable label which allows the polypeptide to be detected. Suitable labels include radioisotopes, e.g. ^{125}I , ^{35}S enzymes, antibodies, polynucleotides and ligands such as biotin. Labelled polypeptides of the invention may be used in diagnostic procedures such as immunoassays in order to determine the amount of a polypeptide of the invention in a sample. Polypeptides or labelled polypeptides of the invention may also be used in serological or cell mediated immune assays for the detection of immune reactivity to said polypeptides in animals and humans using standard protocols.

A polypeptide or labelled polypeptide of the invention or fragment thereof may also be fixed to a solid phase, for example the surface of an immunoassay well, microparticle, dipstick or biosensor. Such labelled and/or immobilized polypeptides may be

packaged into kits in a suitable container along with suitable reagents, controls, instructions and the like.

Such polypeptides and kits may be used in methods of detection of antibodies or cell mediated immunoreactivity, to the
5 mycobacterial proteins and peptides encoded by the ORFs of the invention and their allelic variants and fragments, using immunoassay. Such host antibodies or cell mediated immune reactivity will occur in humans or animals with an immune system which detects and reacts against polypeptides of the invention.
10 The antibodies may be present in a biological sample from such humans or animals, where the biological sample may be a sample as defined above particularly blood, milk or saliva.

Immunoassay methods are well known in the art and will generally comprise:

- 15 (a) providing a polypeptide of the invention comprising an epitope bindable by an antibody against said mycobacterial polypeptide;
- (b) incubating a biological sample with said polypeptide under conditions which allow for the formation of an
20 antibody-antigen complex; and
- (c) determining whether antibody-antigen complex comprising said polypeptide is formed.

Immunoassay methods for cell mediated immune reactivity in animals and humans are also well known in the art (e.g. as
25 described by Weir et al 1994, J.Immunol Methods 176; 93-101) and will generally comprise

- (a) providing a polypeptide of the invention comprising an epitope bindable by a lymphocyte or macrophage or other cell receptor;
- 30 (b) incubating a cell sample with said polypeptide under conditions which allow for a cellular immune response such as release of cytokines or other mediator to occur; and
- (c) detecting the presence of said cytokine or mediator in
35 the incubate.

Polypeptides of the invention may be made by standard synthetic means well known in the art or recombinantly, as described below.

Polypeptides of the invention or fragments thereof labelled or unlabelled may also be used to identify and characterise
5 different strains of *Mptb*, *Mavs*, other GS-containing pathogenic mycobacteria, or *Mtb*, and properties such as drug resistance or susceptibility.

The polypeptides of the invention may conveniently be packaged in the form of a test kit in a suitable container. In such kits
10 the polypeptide may be bound to a solid support where the assay format for which the kit is designed requires such binding. The kit may also contain suitable reagents for treating the sample to be examined, control reagents, instructions, and the like.

The use of polypeptides of the invention in the diagnosis of
15 inflammatory diseases such as Crohn's disease or sarcoidosis in humans or Johne's disease in animals form a preferred aspect of the invention. The polypeptides may also be used in the prognosis of these diseases. For example, the response of a human or animal subject in response to antibiotic or other
20 therapies may be monitored by utilizing the diagnostic methods of the invention over the course of a period of treatment and following such treatment.

The use of *Mtb* polypeptides of the invention in the above-described methods form a further aspect of the invention,
25 particularly for the detection, diagnosis or prognosis of *Mtb* infections.

Polypeptides of the invention may also be used in assay methods for identifying candidate chemical compounds which will be useful in inhibiting, binding to or disrupting the function of said
30 polypeptides required for pathogenicity. In general, such assays involve bringing the polypeptide into contact with a candidate inhibitor compound and observing the ability of the compound to disrupt, bind to or interfere with the polypeptide.

There are a number of ways in which the assay may be formatted. For example, those polypeptides which have an enzymatic function may be assayed using labelled substrates for the enzyme, and the amount of, or rate of, conversion of the substrate into a product
5 measured, e.g by chromatography such as HPLC or by a colourimetric assay. Suitable labels include ^{35}S , ^{125}I , biotin or enzymes such as horse radish peroxidase.

For example, the gene product of ORF C is believed to have GDP-mannose dehydratase activity. Thus an assay for inhibitors of the
10 gene product may utilise for example labelled GDP-mannose, GDP or mannose and the activity of the gene product followed. ORF D encodes a gene related to the synthesis and regulation of capsular polysaccharides, which are often associated with invasiveness and pathogenicity. Labelled polysaccharide
15 substrates may be used in assays of the ORF D gene product. The gene product of ORF F encodes a protein with putative glucosyl transferase activity and thus labelled amino sugars such as β -1-3-N-acetylglucosamine may be used as substrates in assays.

Candidate chemical compounds which may be used may be natural or
20 synthetic chemical compounds used in drug screening programmes. Extracts of plants which contain several characterised or uncharacterised components may also be used.

Alternatively, the a polypeptide of the invention may be screened against a panel of peptides, nucleic acids or other chemical
25 functionalities which are generated by combinatorial chemistry. This will allow the definition of chemical entities which bind to polypeptides of the invention. Typically, the polypeptide of the invention will be brought into contact with a panel of compounds from a combinatorial library, with either the panel
30 or the polypeptide being immobilized on a solid phase, under conditions suitable for the polypeptide to bind to the panel. The solid phase will then be washed under conditions in which only specific interactions between the polypeptide and individual members of the panel are retained, and those specific members may
35 be utilized in further assays or used to design further panels of candidate compounds.

For example, a number of assay methods to define peptide interaction with peptides are known. For example, WO86/00991 describes a method for determining mimotopes which comprises making panels of catamer preparations, for example octamers of
5 amino acids, at which one or more of the positions is defined and the remaining positions are randomly made up of other amino acids, determining which catamer binds to a protein of interest and re-screening the protein of interest against a further panel based on the most reactive catamer in which one or more
10 additional designated positions are systematically varied. This may be repeated throughout a number of cycles and used to build up a sequence of a binding candidate compound of interest.

WO89/03430 describes screening methods which permit the preparation of specific mimotopes which mimic the immunological
15 activity of a desired analyte. These mimotopes are identified by reacting a panel of individual peptides wherein said peptides are of systematically varying hydrophobicity, amphipathic characteristics and charge patterns, using an antibody against an antigen of interest. Thus in the present case antibodies
20 against the a polypeptide of the inventoin may be employed and mimotope peptides from such panels may be identified.

C. Vectors.

Polynucleotides of the invention can be incorporated into a recombinant replicable vector. The vector may be used to
25 replicate the nucleic acid in a compatible host cell. Thus in a further embodiment, the invention provides a method of making polynucleotides of the invention by introducing a polynucleotide of the invention into a replicable vector, introducing the vector into a compatible host cell, and growing the host cell under
30 conditions which bring about replication of the vector. The vector may be recovered from the host cell. Suitable host cells are described below in connection with expression vectors.

D. Expression Vectors.

- 22 -

Preferably, a polynucleotide of the invention in a vector is operably linked to a control sequence which is capable of providing for the expression of the coding sequence by the host cell, i.e. the vector is an expression vector. The term "operably
5 linked" refers to a juxtaposition wherein the components described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the
10 control sequences. Such vectors may be transformed into a suitable host cell as described above to provide for expression of a polypeptide of the invention. Thus, in a further aspect the invention provides a process for preparing polypeptides according to the invention which comprises cultivating a host cell
15 transformed or transfected with an expression vector as described above, under conditions to provide for expression by the vector of a coding sequence encoding the polypeptides, and recovering the expressed polypeptides.

A further embodiment of the invention provides vectors for the
20 replication and expression of polynucleotides of the invention, or fragments thereof. The vectors may be for example, plasmid, virus or phage vectors provided with an origin of replication, optionally a promoter for the expression of the said polynucleotide and optionally a regulator of the promoter. The
25 vectors may contain one or more selectable marker genes, for example an ampicillin resistance gene in the case of a bacterial plasmid or a neomycin resistance gene for a mammalian vector. Vectors may be used *in vitro*, for example for the production of RNA or used to transfect or transform a host cell. The vector
30 may also be adapted to be used *in vivo*, for example in a method of naked DNA vaccination or gene therapy. A further embodiment of the invention provides host cells transformed or transfected with the vectors for the replication and expression of polynucleotides of the invention, including the DNA of GS, the
35 open reading frames thereof and other corresponding ORFs particularly ORFs B, C, E and F from *Mtb*. The cells will be chosen to be compatible with the said vector and may for example be bacterial, yeast, insect or mammalian.

Expression vectors are widely available in the art and can be obtained commercially. Mammalian expression vectors may comprise a mammalian or viral promoter. Mammalian promoters include the metallothionien promoter. Viral promoters include promoters from
5 adenovirus, the SV40 large T promoter and retroviral LTR promoters. Promoters compatible with insect cells include the polyhedrin promoter. Yeast promoters include the alcohol dehydrogenase promoter. Bacterial promoters include the β -galactosidase promoter.

10 The expression vectors may also comprise enhancers, and in the case of eukaryotic vectors polyadenylation signal sequence downstream of the coding sequence being expressed.

Polypeptides of the invention may be expressed in suitable host cells, for example bacterial, yeast, plant, insect and mammalian
15 cells, and recovered using standard purification techniques including, for example affinity chromatography, HPLC or other chromatographic separation techniques.

Polynucleotides according to the invention may also be inserted into the vectors described above in an antisense orientation in
20 order to provide for the production of antisense RNA. Antisense RNA or other antisense polynucleotides or ligands may also be produced by synthetic means. Such antisense polynucleotides may be used in a method of controlling the levels of the proteins encoded by the ORFs of the invention in a mycobacterial cell.

25 Polynucleotides of the invention may also be carried by vectors suitable for gene therapy methods. Such gene therapy methods include those designed to provide vaccination against diseases caused by pathogenic mycobacteria or to boost the immune response of a human or animal infected with a pathogenic mycobacteria.

30 For example, Ziegner et al, AIDS, 1995, 9;43-50 describes the use of a replication defective recombinant amphotropic retrovirus to boost the immune response in patients with HIV infection. Such a retrovirus may be modified to carry a polynucleotide encoding a polypeptide or fragment thereof of the invention and the

retrovirus delivered to the cells of a human or animal subject in order to provide an immune response against said polypeptide. The retrovirus may be delivered directly to the patient or may be used to infect cells ex-vivo, e.g. fibroblast cells, which are then introduced into the patient, optionally after being inactivated. The cells are desirably autologous or HLA-matched cells from the human or animal subject.

Gene therapy methods including methods for boosting an immune response to a particular pathogen are disclosed generally in for example WO95/14091, the disclosure of which is incorporated herein by reference. Recombinant viral vectors include retroviral vectors, adenoviral vectors, adeno-associated viral vectors, vaccinia virus vectors, herpes virus vectors and alphavirus vectors. Alpha virus vectors are described in, for example, WO95/07994, the disclosure of which is incorporated herein by reference.

Where direct administration of the recombinant viral vector is contemplated, either in the form of naked nucleic acid or in the form of packaged particles carrying the nucleic acid this may be done by any suitable means, for example oral administration or intravenous injection. From 10^5 to 10^8 c.f.u of virus represents a typical dose, which may be repeated for example weekly over a period of a few months. Administration of autologous or HLA-matched cells infected with the virus may be more convenient in some cases. This will generally be achieved by administering doses, for example from 10^5 to 10^8 cells per dose which may be repeated as described above.

The recombinant viral vector may further comprise nucleic acid capable of expressing an accessory molecule of the immune system designed to increase the immune response. Such a molecule may be for example interferon, particularly interferon gamma, an interleukin, for example IL-1 α , IL-1 β or IL-2, or an HLA class I or II molecule. This may be particularly desirable where the vector is intended for use in the treatment of humans or animals already infected with a mycobacteria and it is desired to boost the immune response.

E. Antibodies.

The invention also provides monoclonal or polyclonal antibodies to polypeptides of the invention or fragments thereof. The invention further provides a process for the production of monoclonal or polyclonal antibodies to polypeptides of the invention. Monoclonal antibodies may be prepared by conventional hybridoma technology using the polypeptides of the invention or peptide fragments thereof, as immunogens. Polyclonal antibodies may also be prepared by conventional means which comprise inoculating a host animal, for example a rat or a rabbit, with a polypeptide of the invention or peptide fragment thereof and recovering immune serum.

In order that such antibodies may be made, the invention also provides polypeptides of the invention or fragments thereof haptened to another polypeptide for use as immunogens in animals or humans.

For the purposes of this invention, the term "antibody", unless specified to the contrary, includes fragments of whole antibodies which retain their binding activity for a polypeptide of the invention. Such fragments include Fv, F(ab') and F(ab')₂ fragments, as well as single chain antibodies. Furthermore, the antibodies and fragments thereof may be humanised antibodies, e.g. as described in EP-A-239400.

Antibodies may be used in methods of detecting polypeptides of the invention present in biological samples (where such samples include the human or animal body samples, and environmental samples, mentioned above) by a method which comprises:

- (a) providing an antibody of the invention;
- (b) incubating a biological sample with said antibody under conditions which allow for the formation of an antibody-antigen complex; and
- (c) determining whether antibody-antigen complex comprising said antibody is formed.

Antibodies of the invention may be bound to a solid support for example an immunoassay well, microparticle, dipstick or biosensor and/or packaged into kits in a suitable container along with suitable reagents, controls, instructions and the like.

- 5 Antibodies of the invention may be used in the detection, diagnosis and prognosis of diseases as described above in relation to polypeptides of the invention.

F. Compositions.

- 10 The present invention also provides compositions comprising a polynucleotide or polypeptide of the invention together with a carrier or diluent. Compositions of the invention also include compositions comprising a nucleic acid, particularly and expression vector, of the invention. Compositions further include those carrying a recombinant virus of the invention.
- 15 Such compositions include pharmaceutical compositions in which case the carrier or diluent will be pharmaceutically acceptable.

- Pharmaceutically acceptable carriers or diluents include those used in formulations suitable for inhalation as well as oral, parenteral (e.g. intramuscular or intravenous or transcutaneous)
- 20 administration. The formulations may conveniently be presented in unit dosage form and may be prepared by any of the methods well known in the art of pharmacy. Such methods include the step of bringing into association the active ingredient with the carrier which constitutes one or more accessory ingredients. In
- 25 general the formulations are prepared by uniformly and intimately bringing into association the active ingredient with liquid carriers or finely divided solid carriers or both, and then, if necessary, shaping the product.

- For example, formulations suitable for parenteral administration
- 30 include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation isotonic with the blood of the intended recipient, and aqueous and non-aqueous sterile suspensions which may include suspending agents and thickening

agents, and liposomes or other microparticulate systems which are designed to target the polynucleotide or the polypeptide of the invention to blood components or one or more organs, or to target cells such as M cells of the intestine after oral administration.

5 G. Vaccines.

In another aspect, the invention provides novel vaccines for the prevention and treatment of infections caused by *Mptb*, *Mavs*, other GS-containing pathogenic mycobacteria and *Mtb* in animals and humans. The term "vaccine" as used herein means an agent
10 used to stimulate the immune system of a vertebrate, particularly a warm blooded vertebrate including humans, so as to provide protection against future harm by an organism to which the vaccine is directed or to assist in the eradication of an organism in the treatment of established infection. The immune
15 system will be stimulated by the production of cellular immunity antibodies, desirably neutralizing antibodies, directed to epitopes found on or in a pathogenic mycobacterium which expresses any one of the ORFs of the invention. The antibody so produced may be any of the immunological classes, such as the
20 immunoglobulins A, D, E, G or M. Vaccines which stimulate the production of IgA are interest since this is the principle immunoglobulin produced by the secretory system of warm-blooded animals, and the production of such antibodies will help prevent infection or colonization of the intestinal tract. However an
25 IgM and IgG response will also be desirable for systemic infections such as Crohn's disease or tuberculosis.

Vaccines of the invention include polynucleotides of the invention or fragments thereof in suitable vectors and administered by injection of naked DNA using standard protocols.
30 Polynucleotides of the invention or fragments thereof in suitable vectors for the expression of the polypeptides of the invention may be given by injection, inhalation or by mouth. Suitable vectors include *M.bovis* BCG, *M.smegmatis* or other mycobacteria, *Corynebacteria*, *Salmonella* or other agents according to
35 established protocols.

Polypeptides of the invention or fragments thereof in substantially isolated form may be used as vaccines by injection, inhalation, oral administration or by transcutaneous application according to standard protocols. Adjuvants (such as Iscoms or
5 polylactide-coglycolide encapsulation), cytokines such as IL-12 and other immunomodulators may be used for the selective enhancement of the cell mediated or humoral immunological responses. Vaccination with polynucleotides and/or polypeptides of the invention may be undertaken to increase the susceptibility
10 of pathogenic mycobacteria to antimicrobial agents *in vivo*.

In instances wherein the polypeptide is correctly configured so as to provide the correct epitope, but is too small to be immunogenic, the polypeptide may be linked to a suitable carrier.

A number of techniques for obtaining such linkage are known in
15 the art, including the formation of disulfide linkages using N-succinimidyl-3-(2-pyridylthio) propionate (SPDP) and succinimidyl 4-(N-maleimido-methyl)cyclohexane-1-carboxylate (SMCC) obtained from Pierce Company, Rockford, Illinois, (if the peptide lacks a sulfhydryl group, this can be provided by addition of a
20 cysteine residue). These reagents create a disulfide linkage between themselves and peptide cysteine residues on one protein and an amide linkage through the epsilon-amino on a lysine, or other free amino group in the other. A variety of such disulfide/amide-forming agents are known. See, for example,
25 Immun Rev (1982) 62:185. Other bifunctional coupling agents form a thioether rather than a disulfide linkage. Many of these thioether-forming agents are commercially available and include reactive esters of 6-maleimidocaproic acid, 2-bromoacetic acid, 2-iodoacetic acid, 4-(N-maleimido-methyl)cyclohexane-1-carboxylic
30 acid, and the like. The carboxyl group can be activated by combining them with succinimide or 1-hydroxyl-2-nitro-4-sulfonic acid, sodium salt. Additional methods of coupling antigens employs the rotavirus/"binding peptide" system described in EPO Pub. No. 259,149, the disclosure of which is incorporated herein
35 by reference. The foregoing list is not meant to be exhaustive, and modifications of the named compounds can clearly be used.

Any carrier may be used which does not itself induce the production of antibodies harmful to the host. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins; polysaccharides, such as latex functionalized
5 Sepharose®, agarose, cellulose, cellulose beads and the like; polymeric amino acids, such as polyglutamic acid, polylysine, polylactide-coglycolide and the like; amino acid copolymers; and inactive virus particles. Especially useful protein substrates are serum albumins, keyhole limpet hemocyanin, immunoglobulin
10 molecules, thyroglobulin, ovalbumin, tetanus toxoid, and other proteins well known to those skilled in the art.

The immunogenicity of the epitopes may also be enhanced by preparing them in mammalian or yeast systems fused with or assembled with particle-forming proteins such as, for example,
15 that associated with hepatitis B surface antigen. See, e.g., US-A-4,722,840. Constructs wherein the epitope is linked directly to the particle-forming protein coding sequences produce hybrids which are immunogenic with respect to the epitope. In addition, all of the vectors prepared include epitopes specific to HBV,
20 having various degrees of immunogenicity, such as, for example, the pre-S peptide.

In addition, portions of the particle-forming protein coding sequence may be replaced with codons encoding an epitope of the invention. In this replacement, regions which are not required
25 to mediate the aggregation of the units to form immunogenic particles in yeast or mammals can be deleted, thus eliminating additional HBV antigenic sites from competition with the epitope of the invention.

Vaccines may be prepared from one or more immunogenic
30 polypeptides of the invention. These polypeptides may be expressed in various host cells (e.g., bacteria, yeast, insect, or mammalian cells), or alternatively may be isolated from viral preparations or made synthetically.

In addition to the above, it is also possible to prepare live
35 vaccines of attenuated microorganisms which express one or more

recombinant polypeptides of the invention. Suitable attenuated microorganisms are known in the art and include, for example, viruses (e.g., vaccinia virus), as well as bacteria.

5 The preparation of vaccines which contain an immunogenic polypeptide(s) as active ingredients, is known to one skilled in the art. Typically, such vaccines are prepared as injectables, or as suitably encapsulated oral preparations and either liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid prior to ingestion or injection may also
10 be prepared. The preparation may also be emulsified, or the protein encapsulated in liposomes. The active immunogenic ingredients are often mixed with excipients which are pharmaceutically acceptable and compatible with the active ingredient. Suitable excipients are, for example, water, saline,
15 dextrose, glycerol, ethanol, or the like and combinations thereof. In addition, if desired, the vaccine may contain minor amounts of auxiliary substances such as wetting or emulsifying agents, pH buffering agents, and/or adjuvants which enhance the effectiveness of the vaccine. Examples of adjuvants which may
20 be effective include but are not limited to: aluminum hydroxide, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-nor-muramyl-L-alanyl-D-isoglutamine (CGP 11637, referred to as nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine
25 (CGP 19835A, referred to as MTP-PE), and RIBI, which contains three components extracted from bacteria, monophosphoryl lipid A, trehalose dimycolate and cell wall skeleton (MPL+TDM+CWS) in a 2% squalene/Tween® 80 emulsion. The effectiveness of an adjuvant may be determined by measuring the amount of antibodies
30 directed against an immunogenic polypeptide containing an antigenic sequence resulting from administration of this polypeptide in vaccines which are also comprised of the various adjuvants.

The vaccines are conventionally administered parenterally, by
35 injection, for example, either subcutaneously or intramuscularly. Additional formulations which are suitable for other modes of administration include suppositories, oral formulations or as

enemas. For suppositories, traditional binders and carriers may include, for example, polyalkylene glycols or triglycerides; such suppositories may be formed from mixtures containing the active ingredient in the range of 0.5% to 10%, preferably 1% - 2%. Oral
5 formulations include such normally employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like. These compositions take the form of solutions, suspensions, tablets, pills, capsules, sustained
10 release formulations or powders and contain 10% - 95% of active ingredient, preferably 25% - 70%.

The proteins may be formulated into the vaccine as neutral or salt forms. Pharmaceutically acceptable salts include the acid addition salts (formed with free amino groups of the peptide) and
15 which are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids such as acetic, oxalic, tartaric, maleic, and the like. Salts formed with the free carboxyl groups may also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium,
20 or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, 2-ethylamino ethanol, histidine, procaine, and the like.

The vaccines are administered in a manner compatible with the dosage formulation, and in such amount as will be
25 prophylactically and/or therapeutically effective. The quantity to be administered, which is generally in the range of 5 μ g to 250 μ g, of antigen per dose, depends on the subject to be treated, capacity of the subject's immune system to synthesize antibodies, mode of administration and the degree of protection desired.
30 Precise amounts of active ingredient required to be administered may depend on the judgement of the practitioner and may be peculiar to each subject.

The vaccine may be given in a single dose schedule, or preferably in a multiple dose schedule. A multiple dose schedule is one in
35 which a primary course of vaccination may be with 1-10 separate doses, followed by other doses given at subsequent time intervals

required to maintain and or reenforce the immune response, for example, at 1-4 months for a second dose, and if needed, a subsequent dose(s) after several months. The dosage regimen will also, at least in part, be determined by the need of the individual and be dependent upon the judgement of the practitioner.

In a further aspect of the invention, there is provided an attenuated vaccine comprising a normally pathogenic mycobacteria which harbours an attenuating mutation in any one of the genes encoding a polypeptide of the invention. The gene is selected from the group of ORFs A, B, C, D, E, F, G and H, including the homologous ORFs B, C, E and F in *Mtb*.

The mycobacteria may be used in the form of killed bacteria or as a live attenuated vaccine. There are advantages to a live attenuated vaccine. The whole live organism is used, rather than dead cells or selected cell components which may exhibit modified or denatured antigens. Protein antigens in the outer membrane will maintain their tertiary and quaternary structures. Therefore the potential to elicit a good protective long term immunity should be higher.

The term "mutation" and the like refers to a genetic lesion in a gene which renders the gene non-functional. This may be at either the level of transcription or translation. The term thus envisages deletion of the entire gene or substantial portions thereof, and also point mutations in the coding sequence which result in truncated gene products unable to carry out the normal function of the gene.

A mutation introduced into a bacterium of the invention will generally be a non-reverting attenuating mutation. Non-reverting means that for practical purposes the probability of the mutated gene being restored to its normal function is small, for example less than 1 in 10^6 such as less than 1 in 10^9 or even less than 1 in 10^{12} .

- 33 -

An attenuated mycobacteria of the invention may be in isolated form. This is usually desirable when the bacterium is to be used for the purposes of vaccination. The term "isolated" means that the bacterium is in a form in which it can be cultured, processed or otherwise used in a form in which it can be readily identified and in which it is substantially uncontaminated by other bacterial strains, for example non-attenuated parent strains or unrelated bacterial strains. The term "isolated bacterium" thus encompasses cultures of a bacterial mutant of the invention, for example in the form of colonies on a solid medium or in the form of a liquid culture, as well as frozen or dried preparations of the strains.

In a preferred aspect, the attenuated mycobacterium further comprises at least one additional mutation. This may be a mutation in a gene responsible for the production of products essential to bacterial growth which are absent in a human or animal host. For example, mutations to the gene for aspartate semi-aldehyde dehydrogenase (*asd*) have been proposed for the production of attenuated strains of *Salmonella*. The *asd* gene is described further in Gene (1993) 129; 123-128. A lesion in the *asd* gene, encoding the enzyme aspartate β -semialdehyde dehydrogenase would render the organism auxotrophic for the essential nutrient diaminopellic acid (DAP), which can be provided exogenously during bulk culture of the vaccine strain. Since this compound is an essential constituent of the cell wall for gram-negative and some gram-positive organisms and is absent from mammalian or other vertebrate tissues, mutants would undergo lysis after about three rounds of division in such tissues. Analogous mutations may be made to the attenuated mycobacteria of the invention.

In addition or in the alternative, the attenuated mycobacteria may carry a *recA* mutation. The *recA* mutation knocks out homologous recombination - the process which is exploited for the construction of the mutations. Once the *recA* mutation has been incorporated the strain will be unable to repair the constructed deletion mutations. Such a mutation will provide attenuated strains in which the possibility of homologous recombination to

with DNA from wild-type strains has been minimized. RecA genes have been widely studied in the art and their sequences are available. Further modifications may be made for additional safety.

- 5 The invention further provides a process for preparing a vaccine composition comprising an attenuated bacterium according to the invention process comprises (a) inoculating a culture vessel containing a nutrient medium suitable for growth of said bacterium; (b) culturing said bacterium; (c) recovering said
10 bacteria and (d) mixing said bacteria with a pharmaceutically acceptable diluent or carrier.

Attenuated bacterial strains according to the invention may be constructed using recombinant DNA methodology which is known per se. In general, bacterial genes may be mutated by a process of
15 targeted homologous recombination in which a DNA construct containing a mutated form of the gene is introduced into a host bacterium which it is desired to attenuate. The construct will recombine with the wild-type gene carried by the host and thus the mutated gene may be incorporated into the host genome to
20 provide a bacterium of the present invention which may then be isolated.

The mutated gene may be obtained by introducing deletions into the gene, e.g by digesting with a restriction enzyme which cuts the coding sequence twice to excise a portion of the gene and
25 then religating under conditions in which the excised portion is not reintroduced into the cut gene. Alternatively frame shift mutations may be introduced by cutting with a restriction enzyme which leaves overhanging 5' and 3' termini, filling in and/or trimming back the overhangs, and religating. Similar mutations
30 may be made by site directed mutagenesis. These are only examples of the types of techniques which will readily be at the disposal of those of skill in the art.

Various assays are available to detect successful recombination. In the case of attenuations which mutate a target gene necessary
35 for the production of an essential metabolite or catabolite

compound, selection may be carried out by screening for bacteria unable to grow in the absence of such a compound. Bacteria may also be screened with antibodies or nucleic acids of the invention to determine the absence of production of a mutated
5 gene product of the invention or to confirm that the genetic lesion introduced - e.g. a deletion - has been incorporated into the genome of the attenuated strain.

The concentration of the attenuated strain in the vaccine will be formulated to allow convenient unit dosage forms to be
10 prepared. Concentrations of from about 10^4 to 10^9 bacteria per ml will generally be suitable, e.g. from about 10^5 to 10^8 such as about 10^6 per ml. Live attenuated organisms may be administered subcutaneously or intramuscularly at up to 10^8 organisms in one or more doses, e.g. from around 10^5 to 10^8 , e.g. about 10^6 or 10^7
15 organisms in a single dose.

The vaccines of the invention may be administered to recipients to treat established disease or in order to protect them against diseases caused by the corresponding wild type mycobacteria, such as inflammatory diseases such as Crohn's disease or sarcoidosis
20 in humans or Johne's disease in animals. The vaccine may be administered by any suitable route. In general, subcutaneous or intramuscular injection is most convenient, but oral, intranasal and colorectal administration may also be used.

The following Examples illustrates aspects of the invention.

25 EXAMPLE 1

Tests for the presence of the GS identifier sequence were performed on $5\mu\text{l}$ bacterial DNA extracts ($25\mu\text{g/ml}$ to $500\mu\text{g/ml}$) using polymerase chain reaction based on the oligonucleotide
30 primers 5'-GATGCCGTGAGGAGGTAAAGCTGC-3' (Seq ID No. 40) and 5'-GATACGGCTCTTGAATCCTGCACG-3' (Seq ID No. 41) from within the identifier DNA sequences (Seq. ID Nos 1 and 2). PCR was performed for 40 cycles in the presence of 1.5 mM magnesium and an annealing temperature of 58°C . The presence or absence of the correct amplification product indicated the presence or absence

of GS identifier sequence in the corresponding bacterium. GS identifier sequence is shown to be present in all the laboratory and field strains of *Mptb* and *Mavs* tested. This includes *Mptb* isolates 0025 (bovine CVL Weybridge), 0021 (caprine, Moredun), 5 0022 (bovine, Moredun), 0139 (human, Chiodini 1984), 0209, 0208, 0211, 0210, 0212, 0207, 0204, 0206 (bovine, Whipple 1990). All *Mptb* strains were IS900 positive. The *Mavs* strains include 0010 and 0012 (woodpigeon, Thorel) 0018 (armadillo, Portaels) and 0034, 0037, 0038, 0040 (AIDS, Hoffner). All *Mavs* strains were 10 IS902 positive. One pathogenic *M.avium* strain 0033 (AIDS, Hoffner) also contained GS identifier sequence. GS identifier sequence is absent from other mycobacteria including other *M.avium*, *M.malmoense*, *M.szulgai*, *M.gordonae*, *M.chelonae*, *M.fortuitum*, *M.phlei*, as well as *E.coli*, *S.areus*, *Nocardia* sp, 15 *Streptococcus* sp. *Shigella* sp. *Pseudomonas* sp.

Example 2:

To obtain the full sequence of GS in *Mavs* and *Mptb* we generated a genomic library of *Mavs* using the restriction endonuclease EcoRI and cloning into the vector pUC18. This achieved a 20 representative library which was screened with ³²P-labelled identifier sequence yielding a positive clone containing a 17kbp insert. We constructed a restriction map of this insert and identified GS as fragments unique to *Mavs* and *Mptb* and not occurring in laboratory strains of *M.avium*. These fragments 25 were sub-cloned into pUC18 and pGEM4Z. We identified GS contained within an 8kb region. The full nucleotide sequence was determined for GS on both DNA strands using primer walking and automated DNA sequencing. DNA sequence for GS in *Mptb* was obtained using overlapping PCR products generated using PwoDNA 30 polymerase, a proofreading thermostable enzyme. The final DNA sequences were derived using the University of Wisconsin GCG gel assembly software package.

Example 3:

The DNA sequence of GS in *Mavs* and *Mptb* was found to be more 35 than 99% homologous. The ORFs encoded in GS were identified using GeneRunner and DNASTar computer programmes. Eight ORFs were identified and designated GSA, GSB, GSC, GSD, GSE, GSF, GSG

and GSH. Database comparisons were carried out against the GenEMBL Database release version 48.0 (9/96), using the BLAST and BLIXEM programmes. GSA and GSB encoded proteins of 13.5kDa and 30.7kDa respectively, both of unknown functions. GSC encoded a protein of 38.4kDa with a 65% homology to the amino acid sequence of *rfbD* of *V.cholerae*, a 62% amino acid sequence homology to *gmd* of *E.coli* and a 58% homology to *gca* of *Ps.aeruginosa* which are all GDP-D-mannose dehydratases. Equivalent gene products in *H.influenzae*, *S.dysenteriae*, *Y.enterocolitica*, *N.gonorrhoea*, *K.pneumoniae* and *rfbD* in *Salmonella enterica* are all involved in 'O'-antigen processing known to be linked to pathogenicity. GSD encoded a protein of 37.1kDa which showed 58% homology at the DNA level to *wcaG* from *E.coli*, a gene involved in the synthesis and regulation of capsular polysaccharides, also related to pathogenicity. GSE was found to have a > 30% amino acid homology to *rfbT* of *V.cholerae*, involved in the transport of specific LPS components across the cell membrane. In *V.cholerae* the gene product causes a seroconversion from the Inaba to the Ogawa 'epidemic' strain. GSF encoded a protein of 30.2kDa which was homologous in the range 25-40% at the amino acid level to several glucosyl transferases such as *rfpA* of *K.pneumoniae*, *rfbB* of *K.pneumoniae*, *lgtD* of *H.influenzae*, *lsi* of *N.gonorrhoea*. In *E.coli* an equivalent gene *galE* adds β -1-3 N-acetylglucosamine to galactose, the latter only found in 'O' and 'M' antigens which are also related to pathogenicity. GSH comprising the ORFs GSH₁ and GSH₂ encodes a protein totalling about 60kDa which is a putative transposase with a 40 - 43% homology at the amino acid level to the equivalent gene product of IS21 in *E.coli*. This family of insertion sequences is broadly distributed amongst gram negative bacteria and is responsible for mobility and transposition of genetic elements. An IS21- like element in *B.fragilis* is split either side of the β -lactamase gene controlling its activation and expression. We programmed an *E.coli* S30 cell-free extract with plasmid DNA containing the ORF GSH under the control of a lac promoter in the presence of a ³⁵S-methionine, and demonstrated the translation of an abundant 60kDa protein. The proteins homologous to GS encoded in other organisms are in general highly antigenic. Thus the proteins encoded by the ORFs

in GS may be used in immunoassays of antibody or cell mediated immuno-reactivity for diagnosing infections caused by mycobacteria, particularly *Mptb*, *Mavs* and *Mtb*. Enhancement of host immune recognition of GS encoded proteins by vaccination
5 using naked specific DNA or recombinant GS proteins, may be used in the prevention and treatment of infections caused by *Mptb*, *Mavs* and *Mtb* in humans and animals. Mutation or deletion of all or some of the ORFs A to H in GS may be used to generate attenuated strains of *Mptb*, *Mavs* or *Mtb* with lower pathogenicity
10 for use as living or killed vaccines in humans and animals. Such vaccines are particularly relevant to Johne's disease in animals, to diseases caused by *Mptb* in humans such as Crohn's disease, and to the management of tuberculosis especially where the disease is caused by multiple drug-resistant organisms.

SEQUENCE LISTING

Seq. ID No.1

5' - 1 GATCCAAC TA AACCCGATGG AACCCCGCGC AAACCTATTGG ACGTCTCCGC GCTACGCAGT
 61 TGGGTTGGCG CCCGCGAATC GCACTGAAAG AGGGCATCGA TGCAACGGTG TCGTGGTACC
 121 GCACAAATGC CGATGCCGTG AGGAGGTAAA GCTGCGGGCC GGCCGATGTT ATCCCTCCGG
 181 CCGGACGGGT AGGGCGACCT GCCATCGAGT GGTACGGCAG TCGCCTGGCC GGCAGGGCGC
 241 ATGGCCTATG TGAGTATCCC ATAGCCTGGC TTGGCTCGCC CCTACGCATT ATCAGTTGAC
 301 CGCTTTTCGG CCACGTCGCA GGCTTGGCGC AGCATCCCGT TCAGGTCTCC TCATGGTCCG
 361 GTGTGGCAGC ACCACGCAAG CTCGAACCGA CTCGTTTCCC AATTTCCGAT GCTAATATCG
 421 CTCGATGGAT TTTTTCGCA ACGCCGGCTT GATGGCTCGT AACGTTAGCA CCGAGATGCT
 481 GCGCCACTCC GAACGAAAGC GCCTATTAGT AAACCAAGTC GAAGCATACG GAGTCAACGT
 541 TGTTATTGAT GTCGGTGCTA ACTCCGGCCA GTTCGGTAGC GCTTTGCGTC GTGCAGGATT
 601 CAAGAGCCGT ATCGTTTCCT TTGAACCTCT TTCGGGGCCA TTTGCGCAAC TAACGCGCAA
 661 GTCGGCATCG GATC -3'

15 Seq. ID No.2

5' - 1 GATCCGATGC CGACTTGCGC GTTAGTTGCG CAAATGGCCC CGAAAGAGGT TCAAAGGAAA
 61 CGATACGGCT CTTGAATCCT GCACGACGCA AAGCGCTACC GAACTGGCCG GAGTTAGCAC
 121 CGACATCAAT AACAACTTTG ACTCCGTATG CTTGCACTTG GTTTACTAAT AGGCGCTTTC
 181 GTTCGGAGTG GCGCAGCATC TCGGTGCTAA CGTTACGAGC CATCAAGCCG GCGTTGCGCA
 241 AAAAATCCAT CGAGCGATAT TAGCATGCGA AATTGGGAAA CGAGTCGGTT CGAGCTTGCG
 301 TGGTCGTGCC ACACCGGACC ATGAGGAGAC CTGAACGGGA TGCTGCCGCA AGCCTGCGAC
 361 GTGGCGCGAA AGCGGTCAAC TGATAATGCG TAGGGGGCGAG CCAAGCCAGG CTATGGGATA
 421 CTCACATAGG CCATGCGCCT CGCCGGCCAG GCGACTGCCG TACCACTCGA TGGCAGGTGG
 481 CCCTACCCGT CCGGCCGGAG GGATAACATC GGCGGGCCG CAGCTTTACC TCCTCACGGC
 541 ATCGGCATTT GTGCGGTACC ACGACACCGT TGCATCGATG CCCTCTTTCA GTGCGATTGG
 601 CCGGCGCCAA CCCAACTGCG TAGCGCGGAG ACGTCCAATA GTTTGCGCGG GGTTCATCG
 661 GGTTTAGTTG GATC -3'

Seq. ID No.3

1 GAATTCCTGGG TTGGAGACGA CGTCGAAC TC GTGGTCGGTC TTGCTTCGAA
 51 TGATCGCTGT GATCTGGTCG GCGGTGCCGA CAGGAACCGT CGACTTGTCTG
 101 ACGATCACCT TGTACCGGTC GATGTATGAC CCAATGTCGT CCGCAACCGA
 5 151 GAAGACGTAC GTCAGGTCCG CCGCCCCGCT TTCACCCATG GCGGTCCGGA
 201 CGGCGATGAA AATGACGTCC GCGTGCTCGA TTCGCGTTG CCGGTCCGTG
 251 GTGAAGTCAA TCAGCCCGTT CTCACGGTTC CTCGCAATCA ACTCCCAACC
 301 CGGGCTCGAA AATCGGGACA CTGCCTGCGA GGAGCAAATC GATCTTGGCC
 351 TGATCGATAT CGACACAGAC GACATCGTTG CCGCTATCCG CGAGACAGGC
 10 401 GCGCGTGACG AGGCCTACAT AGCCTGATCC GACCACCGAA ATTTTCAAGA
 451 TGACCCCTTC AAGTCCCGCA TCGGTGACG ACCATACTGC CGCAACTCTG
 501 TACCCTCCGT GGGTAATTCG CATGTGCGGT TCGTAAGGAG CAGCCAGCGA
 551 GTCGGGGACG TTCGGTGAGA GAGTCGCAGG ACTACGAGGT TGCCGGTGCG
 601 ATACATCACA GTGTTGGGTC TGTCGGCAAC GATGCAGCAA GAACCCACGG
 15 651 GGCAGCCCTG AACTGCGCGC ATGACCGGTC CTGTCTCTGG CACCTTTGAT
 701 CGGCCACCGC TTCCATGCGA ACATGACCGG AATCCATAGC GCGTGGTCAA
 751 GCAGCGGGGA GGTAGACGTC GGTGTCTCT GCTCCAACCG TGTCGGTGAT
 801 AACGATTTTC CTGAACGATC TCGAGGGATT GAAAAGCACC GTGGAGAGCG
 851 TTCGCGCGCA GCGCTATGGG GGGCGAATCG AGCACATCGT CATCGACGGT
 20 901 GGATCGGGCG ACGCCGTCGT GGAGTATCTG TCCGGCGATC CTGGCTTTGC
 951 ATATTGGCAA TCTCAGCCCG ACAACGGGAG ATATGACGCG ATGAATCAGG
 1001 GCATTGCCCA TTCGTCCGGC GACCTGTTGT GGTTTATGCA CTCCACGGAT
 1051 CGTTTCTCCG ATCCAGATGC AGTCGCTTCC GTGGTGGAGG CGCTCTCGGG
 1101 GCATGGACCA GTACGTGATT TGTGGGGTTA CGGGAAAAAC AACCTTGTCTG
 25 1151 GACTCGACGG CAAACCACTT TTCCCTCGGC CGTACGGCTA TATGCCGTTT
 1201 AAGATGCGGA AATTTCTGCT CCGCGCGACG GTTGCGCATC AGGCGACATT
 1251 CTTGCGGCGC TCGCTGGTAG CCAAGTTGGG CGGTTACGAT CTTGATTTTG
 1301 GACTCGAGGC GGACCAGCTG TTCATCTACC GTGCCGCACT AATACGGCCT
 1351 CCCGTACGA TCGACCGCGT GGTTCGCGAC TTCGATGTCA CCGGACCTGG
 30 1401 TTCAACCCAG CCCATCCGTG AGCACTATCG GACCCTGCGG CGGCTCTGGG
 1451 ACCTGCATGG CGACTACCGG CTGGGTGGGC GCAGAGTGTG GTGGGCTTAC
 1501 TTGCGTGTGA AGGAGTACTT GATTCGGGCG GACCTGGCGG CATTCAACGC
 1551 GGTAAGTTC TTGCGAGCGA AGTTCGCCAG AGCTTCGCGG AAGCAAAATT
 1601 CATAGAAACC AACTTCTACT GCCTGACCTG AGCAGCGCGG AGGCGCGCAG
 35 1651 CCGGATCAGT GCGACCTGAA CCGCCAGGTG GAAAGCGCCA CCGATCCCGG
 1701 CACCGAGTGC CTGACGCTTC GGATCCCTTG CACCACAACG AGAGTGAGAG
 1751 CGCCATGATG AGGAAATATC GGCTGGGCGG AGTCAACGCC GGAGTGACAA
 1801 AAGTGAGAAC CCGGTGAAGC GAGCGCTTAT AACAGGGATC ACGGGGCGAG
 1851 ATGGTTCCCTA CCTCGCCGAG CTACTACTGA GCAAGGGATA CGAGGTTTAC
 40 1901 GGGCTCGTTC GTCGAGCTTC GACGTTTAA ACCTCGCGGA TCGATCACCT
 1951 CTACGTTGAC CCACACCAAC CCGGCGCGCG CTTGTTCTTG CACTATGCAG
 2001 ACCTCACTGA CCGCACCCGG TTGGTGACCC TGCTCAGCAG TATCGACCCG
 2051 GATGAGGTCT ACAACCTCGC AGCGCAGTCC CATGTGCGCG TCAGCTTTGA
 2101 CGAGCCAGTG CATACCGGAG ACACCACCGG CATGGGATCG ATCCGACTTC
 45 2151 TGGAAGCAGT CCGCTTTTCT CCGGTGGACT GCCGTTCTA TCAGGCTTCC
 2201 TCGTCGGAGA TGTTCCGGCG ATCTCCGCCA CCGCAGAACG AATCGACGCC
 2251 GTTCTATCCC CGTTCCGCAT ACGGCGCGGC CAAGGTCTTC TCGTACTGGA
 2301 CGACTCGCAA CTATCGAGAG GCGTACGGAT TATTGCGAGT GAATGGCATC
 2351 TTGTTCAACC ATGAGTCCCC CCGGCGCGGC GAGACTTTTC TGACCCGAAA
 50 2401 GATCAGCGGT GCCGTGGCGC GCATCCGAGC TGGCGTCCAA TCGGAGGTCT
 2451 ATATGGGCAA CCTCGATGCG ATCCGCGACT GGGGCTACGC GCCCGAATAT
 2501 GTCGAGGGGA TGTGAGGAT GTTGCAAGCG CCTGAACCTG ATGACTACGT

2551 CCTGGCGACA GGGCGTGGTT ACACCGTACG TGAGTTCGCT CAAGCTGCTT
 2601 TTGACCATGT CCGGCTCGAC TGGCAAAAGC GCGTCAAGTT TGACGACCGC
 2651 TATTTGCGTC CCACCGAGGT CGATTGCTA GTAGGAGATG CCGACAAGGC
 2701 GGGCCAGTCA CTCGGCTGGA AAGCTTCGGT TCATACTGGT GAACTCGCGC
 5 2751 GCATCATGGT GGACGCGGAC ATCGCGCGGT TGGAGTGCGA TGGCACACCA
 2801 TGGATCGACA CGCCGATGTT GCCTGGTTGG GGCAGAGTAA GTTGACGACT
 2851 ACACCTGGGC CTCTGGACCG CGCAACGCCC GTGTATATCG CCGGTCATCG
 2901 GGGGCTGGTC GGCTCAGCGC TCGTACGTAG ATTTGAGGCC GAGGGGTTCA
 2951 CCAATCTCAT TGTGCGATCA CGCGATGAGA TTGATCTGAC GGACCGAGCC
 10 3001 GCAACGTTTG ATTTGTGTG TGAGACAAGA CCACAGGTGA TCATCGATGC
 3051 GGCCGCACGG GTCGCGGGCA TCATGGCGAA TAACACCTAT CCCGCGGACT
 3101 TCTTGTCCGA AAACCTCCGA ATCCAGACCA ATTTGCTCGA CGCAGCTGTC
 3151 GCGGTGCGTG TGCCGCGGCT CCTTTCTCTC GGTTCGTCTAT GCATCTACCC
 3201 GAAGTACGCT CCGCAACCTA TCCACGAGAG TGCTTTATTG ACTGGCCCTT
 15 3251 TGGAGCCAC CAACGACGCG TATGCGATCG CCAAGATCGC CCGTATCCTG
 3301 CAAGTTCAGG CGGTTAGGCG CCAATATGGG CTGGCGTGA TCTCTGCGAT
 3351 GCGGACTAAC CTCTACGGAC CCGGCGACAA CTCTCCCCG TCCGGGTGCG
 3401 ATCTCTTGCC GCGGCTCATC CGTCGATATG AGGAAAGCAA AGCTGGTGGT
 3451 GCAGAAGAGG TGACGAATTG GGGGACCGGT ACTCCGCGGC GCGAACTTCT
 20 3501 GCATGTCGAC GATCTGCGA GCGCATGCCT GTTCTTTTG GAACATTTCG
 3551 ATGGTCCGAA CCACGTCAAC GTGGGACCG GCGTCGATCA CAGCATTAGC
 3601 GAGATCGCAG ACATGGTCGC TACAGCGGTG GGCTACATCG GCGAAACACG
 3651 TTGGGATCCA ACTAAACCCG ATGGAACCCC GCGCAAACCTA TTGGACGTCT
 3701 CCGCGCTACG CGAGTTGGGT TGGCGCCCGC GAATCGCACT GAAAGACGGC
 25 3751 ATCGATGCAA CCGTGTCTGT GTACCGCACA AATGCCGATG CCGTGAGGAG
 3801 GTAAAGCTGC GGGTCGGCCG ATGTTATCCC TCCGGCCGGA CCGGTGGGGC
 3851 GACCTGCCGT CGAGTGGTAC GGCAGTCGCC TGGCCGCGA GCGCGGTGGC
 3901 CTATGGGAGT ATCCAATAGC CTGGCTTGGC TCGCCCTAC GCATTATCAG
 3951 TTGACCGCTT TCGCGCCAGC TCGCAGGCTT GCGGCAGCAT CCCGTTGAGG
 30 4001 TCTCCTCATG GTCCGTTGTG GCACGACCAC GCAAGCTCGA ACCGACTCGT
 4051 TTCCCAATTT CGCATGCTAA TATCGCTCGA TGGATTTTTT GCGCAACGCC
 4101 GGCTTGATGG CTCGTAACGT TAGTACCGAG ATGCTGCCCC ACTTCGAACG
 4151 AAAGCGCCTA TTAGTAAACC AATTCAAAGC ATACGGAGTC AACGTTGTTA
 4201 TTGATGTCTG TGCTAACTCC GGCCAGTTCT GTAGCGCTTT GCGTCGTGCA
 35 4251 GGATTCAAGA GCGTATCGT TTCCTTTGAA CCTCTTTGCG GGCCATTGTC
 4301 GCAACTAAGC CGCAAGTCGG CATCGGATCC ACTATGGGAG TGTCACCAGT
 4351 ATGCCCTAGG CGACGCCGAT GAGACGATTA CCATCAATGT GGCAGGCAAT
 4401 GCGGGGGCAA GTAGTTCGCT GCTGCCGATG CTTAAAAGTC ATCAAGATGC
 4451 CTTTCCTCCC GCGAATTATA TTGGCACCAG AGACGTTGCA ATACACCGCC
 40 4501 TTGATTCGGT TGCATCAGAA TTTCTGAACC CTACCGATGT TACTTTCCTG
 4551 AAGATCGACC TACAGGGTTT CGAGAAGCAG GTTATCACGG GCAGTAAGTC
 4601 AACGCTTAAC GAAAGCTGCG TCGGCATGCA ACTCGAACTT TCTTTTATTC
 4651 CGTTGTACGA AGGTGACATG CTGATTCATG AAGCGCTTGA ACTTGTCTAT
 4701 TCCCTAGGTT TCAGACTGAC GGGTTTGTG CCCGGCTTTA CCGATCCGGC
 45 4751 CAATGGTTCGA ATGCTTCAAG CTGACGGCAT TTTCTTCCGT GGGGACGATT
 4801 GACATAAATG CTCCGTCGGC ACCCTGCCGG TATCCAAACG GCGATCTGG
 4851 TGAGCCGGCC TCCCGGGCAC CTAATCGACT ATCTAAATTG AGGCGGCCGC
 4901 GACGTGCGGC ACGAACAGGT GGCCGGCTGC TAGCGTTACA CACGTCATGA
 4951 CTGCGCCAGT GTTCTCGATA ATTATCCCTA CTTTCAATGC AGCGGTGACG
 50 5001 CTGCAAGCCT GCCTCGGAAG CATCGTCGGG CAGACCTACC GGGAAAGTGA
 5051 AGTGGTCCTT GTCGACGGCG GTTCGACCGA TCGGACCTTC GACATCGCGA
 5101 ACAGTTTCCG CCCGGAACCTC GGCTCGCGAC TGGTCTTCA CAGCGGGCCC
 5151 GATGATGGCC CCTACGACGC CATGAACCGC GCGTCGGCG TGGCCACAGG

5201 CGAATGGGTA CTTTTTTTAG GCGCCGACGA CACCCTCTAC GAACCAACCA
 5251 CGTTGGCCCA GGTAGCCGCT TTTCTCGGCG ACCATGCGGC AAGCCATCTT
 5301 GTCTATGGCG ATGTTGTGAT GCGTTCGACG AAAAGCCGGC ATGCCGGACC
 5351 TTTCGACCTC GACCGCCTCC TATTTGAGAC GAATTTGTGC CACCAATCGA
 5401 TCTTTTACCG CCGTGAGCTT TTCGACGGCA TCGGCCCTTA CAACCTGCGC
 5451 TACCGAGTCT GGGCGGACTG GGACTTCAAT ATTCGCTGCT TCTCCAACCC
 5501 GCGCGTGATT ACCCGCTACA TGGACGTCGT GATTTCCGAA TACAACGACA
 5551 TGACCGGCTT CAGCATGAGG CAGGGGACTG ATAAAGAGTT CAGAAAACGG
 5601 CTGCCAATGT ACTTCTGGGT TGCAGGGTGG GAGACTTGCA GGCGCATGCT
 5651 GGCGTTTTTG AAAGACAAGG AGAATCGCCG TCTGGCCTTG CGTACGCGGT
 5701 TGATAAGGGT TAAGGCCGTC TCCAAAGAAC GAAGCGCAGA ACCGTAGTCG
 5751 CGGATCCACA TTGGACTTCT TTAACGCGTT TCGTCTCTGA TCCACCTTTC
 5801 AAGCCCGTTC CGCGTAACGC GCGCGCGAGA GAGTGGTCGC ATATCGCATC
 5851 ACTGTTCTCG TGCCAGTGCT TGGAAAGCGT CGAGCACTCT GGTTCGCGTT
 5901 CTTGACGTTT GCGCCCGCTC CTAGAGGTAG CGTGTCACGT GACTGAAGCC
 5951 AATGAGTGCA ACTCGGCGTC GCGAAAGGTT TCAGTCGCGG TTGAGCAAGA
 6001 CACCGCAAGA CTACTGGAGT GCGTGACAA GCGCCTCCAG CTCGCGGCTG
 6051 AAAGCGGATG CAAAGGGATT CGAAGCTTGA GCAACATGCG AAGGGGAGAA
 6101 CGGCCTATGA GGCTGGGACA GGTTTTCGAT CCGCGCGCGA ATGCACTGTC
 6151 AATGGCCAAG TAGAAGTCCC CGCTGGTGGC CAGCAGAAGT CCCCCTCCG
 6201 CTGCGGGTGG TTGGCTAATT CTTGGCGGCT CCCTTCTTGT GGTGCGGCTG
 6251 GCGCATCCGG TAGGACTCGC CGGAGGTGAC GACGATGCTG CCGTGGTGCA
 6301 GCAGCCGATC GAGGATGCTG GCGGCGGTGG TGTGCTCGGG CAGGAATCGC
 6351 CCCCATTGTT CGAAGGGCCA ATGCGAGGCG ATGGCCAGGG AGCGGCGCTC
 6401 GTAGCCGGCA GCCACGAGCC GGAACAACAG TTGAGTCCCG GTGTCGTCCA
 6451 GCGGGGCGAA GCGGATCTCG TCCAAGATGA CCAGATCCGC GCGGAGCAGG
 6501 GTGTGATGTA TCTTGGCGAC GGTGTTGTG GCGAGGCGCG GGTAGAGGAC
 6551 CTCGATCAGG TCGGCGGCGG TGAAGTAGCG GACTTTGAAT CCGGCGTGGA
 6601 CCGCAGCGTG CCGCAGCCG ATGAGCAGGT GACTTTTGCC CGTACCAGGT
 6651 GGGCCAATGA CCGCCAGGTT CTGTTGTGCC CGAATCCATT CCAGGCTCGA
 6701 CAGGTAGTCG AACGTGGCTG CCGTGATCGA CGATCCGGTG ACGTCGAACC
 6751 CGTGAGGGT CTTGGTGACC GGAAGGGCTG CCGCCTTGAG ACGGTTGGCG
 6801 GTGTTGGAGG CATCGCGGCG AGCGATCTCG GCCTCAACCA ACGTCCGCG
 6851 GATCTCCTCC GGTGTCCAGC GTTGGCTCTT GCGGACTTGC AACACCTCGG
 6901 CCGCGTTGCG GCGCACCGTG GCCAGCTTCA ACCGCGCGAG CGCCGCGTCA
 6951 AGGTCAGCAG CCAGCGGTGC CCGCAGGAC GGTGCCACCG GCTTGGCAGC
 7001 GGTGGTCATG AGGCCGTCCC GTCGGTGGTG TTGATCTTGT AGGCCTCCAA
 7051 CGAGCGGGTC TCGACGGTGG GCAGATCGAG CACGAGTGCG TCGCCGGCGG
 7101 GCGGGGGTTG TGGGGTGCCG GCGCCGGCGG CCAGGATCGA GCGCACGTG
 7151 GCAGCGCGGA ACCGGCGAAA CGCAACCGCC CCGCGCAGCG CGTCAATCAA
 7201 AGCCTGTTCC CCGTGGGCGG CGCCAAGGCC GAGCAGAATG TCGAGTTCGG
 7251 ATTTCACTCG GGTGTTGCCG ATCGCAGCAG CACCGACGAG GAACTGCTGC
 7301 GCTTCGGTTC CCAATGCGCA GAATCGTTTC TCTGCTTGGG TTTTCGGGCG
 7351 AGGACCACGC GAGGGTGCGG GTCTGGGTCC GTCGTAGTGT TCATCGAGGA
 7401 TGGACACCTC ACCTGGGCTG ACGAGCTCGT GTCGCGCCAC GATCACACCG
 7451 GTCGAGGTT CCAACAGGAT CAGGGCGCCA TGATCGACCA CCACCGCCAC
 7501 GGTGGCACCG ACGAGCCGCT GAGGCACCGA GTAACGAGCT GAGCGGTAAC
 7551 GGATGCACGA GAGGCCGTCG ACCTTACGGC GCACCGACCC CGAGCCGATC
 7601 GTCGGCCGCA GCGAGGCGAG CTCCCTCAAG ACGGTGCGCT CGTCAACCAA
 7651 GCGATCGTTG GGCACGCGC AGATCTCCGA GTGGACCGTG GCATTGACCT
 7701 CCGCGCACCA TAGTTGCGCC TGGGCGTTGA GGGCACGTAG GTCGACCTGC
 7751 TCACCGGCTA ACGCAGCTTC GGTGAGCAGC GGCACCGCAA GGTGCTCCTG
 7801 AGCGTAGCCA CAGAGGTTCT CCACGATGCC CTTGATGTC GGATCCGCAC

7851 CGTGGCAGAA GTCCGGAACG AAGCCATAGT GGGACGGGAA TCGCACATAA
 7901 TCCGGTGTGTG GAACAACAAC ATTGGCGACG ACACCACCTT TGAGGCAGCC
 7951 CATCCGGTCG GCCAGGATCT TGGCCGGAAC CCCACCGATC GCCTC

Seq. ID No. 4

5 1 TTCTACTGCC TGACCTGAGC AGCGCCGAGG CGCGCAGCGC GATCACTGCG ACCTGAATGG
 61 CCAGGTGGAA AGCGCCACCG ATCCCGGCAC CGAGTGCCCTG ACGATTCCGA TCCCTTGACAC
 121 CACAACGAGA GTGAGACCGC CATGATGACG AAATATCGGC TGGGCGGAGT CAACGCCGGA
 181 GTGACAAAAG TGAGAACCCG GTGAAGCGAG CGCTTATAAC AGGGATCAGC GGGCAGGATG
 241 GTTCCTACCT CGCCGAGCTA CTA CTGAGCA AGGGATACGA GGTTCACGGG CTCGTTCTGTC
 10 301 GAGCTTCGAC GTTTAACACG TCGCGGATCG ATCACCCTTA CGTTGACCCA CACCAACCGG
 361 GCGCGCGCTT GTTCTTGAC TATGCAGACC TCACTGACGG CACCCGGTTG GTGACCCTGC
 421 TCAGCAGTAT CGACCCGGAT GAGGTCTACA ACCTCGCAGC GCAGTCCCAT GTGCGCGTCA
 481 GCTTTGACGA GCCAGTGCAT ACCGGAGACA CCACCGGCAT GGGATCGATC CGACTTCTGG
 541 AAGCAGTCCG CCTTTCTCGG GTGGACTGCC GGTTCCTATCA GGCTTCTCTG TCGGAGATGT
 15 601 TCGGCGCATC TCCGCCACCG CAGAACGAAT CGACGCCGTT CTATCCCGT TCGCCATACG
 661 GCGCGGCCAA GGTCTTCTCG TACTGGACGA CTCGCAACTA TCGAGAGCGG TACGGATTAT
 721 TCGCAGTGAA TGGCATCTTG TTCAACCATG AGTCCCCCGG GCGCGGCGAG ACTTTCGTGA
 781 CCCGAAAGAT CACGCGTGCC GTGGCGCGCA TCCGAGCTGG CGTCCAATCG GAGGTCTATA
 841 TGGGCAACCT CGATGCGATC CGCGACTGGG GCTACGCGCC CGAATATGTC GAGGGGATGT
 20 901 GGAGGATGTT GCAAGCGCCT GAACCTGATG ACTACGTCTT GCGGACAGGG CGTGGTTACA
 961 CGGTACGTGA GTTCGCTCAA GCTGCTTTTG ACCACGTCCG GCTCGACTGG CAAAAGCAGC
 1021 TCAAGTTTGA CGACCCGTAT TTGCGCCCA CCGAGGTGCA TTCGCTAGTA GGAGATGCCG
 1081 ACAGGGCGGC CCAGTCACTC GGCTGGAAAG CTTCGGTTCA TACTGGTGAA CTCGCGCGCA
 1141 TCATGGTGGA CGCGGACATC GCGCGCTCGG AGTGCGATGG CACACCATGG ATCGACACGC
 25 1201 CGATGTTGCC TGGTTGGGGG GGAGTAAGTT GACGACTACA CCTGGGCCTC TGGACCGCGC
 1261 AACGCCCCGTG TATATCGCCG GTCATCGGGG GCTGGTCCGC TCAGCGCTCG TACGTAGATT
 1321 TGAGGCCGAG GGGTTACCA ATCTCATTGT GCGATCACGC GATGAGATTG ATCTGACGGA
 1381 CCGAGCCGCA ACGTTTGATT TTGTGTCTGA GACAAGACCA CAGGTGATCA TCGATGCCGG
 1441 CGCACGGGTC GCGGGCATCA TGGCGAATAA CACCTATCCC GCGGACTTCT TGTCCGAAAA
 30 1501 CCTCCGAATC CAGACCAATT TGCTCGACGC AGCTGTGCGC GTGCGTGTGC CGCGGCTCCT
 1561 TTTCTCTGGT TCGTCAATGA TCTACCCGAA GTACGCTCCG CAACCTATCC ACGAGAGTGC
 1621 TTTATTGACT GGCCCTTTGG AGCCCAACCA CGACGCTAT GCGATCGCCA AGATCGCCGG
 1681 TATCCTGCAA GTTCAGGCGG TTAGGCGCCA ATATGGGCTG GCGTGGATCT CTGCGATGCC
 1741 GACTAACCTC TACGGACCCG GCGACAACCT CTCGCCGTCC GGGTCGCATC TCTTGCCGGC
 35 1801 GCTCATCCGT CGATATGAGG AAGCCAAAGC TGGTGGTGCA GAAGAGGTGA CGAATTGGGG
 1861 GACCGGTACT CCGCGGCGCG AACTTCTGCA TGTCGACGAT CTGGCGAGCG CATGCCGTGT
 1921 CCTTTTGAA CATTTCGATG GTCCGAACCA CGTCAACGTG GGCACCGGCG TCGATCACAG
 1981 CATTAGCGAG ATCGCAGACA TGGTCCGTAC GCGGGTGGG TACATCGGCG AAACACGTTG
 2041 GGATCCAACT AAACCCGATG GAACCCCGCG CAACTATFG GACGTCTCCG CGCTACGCGA
 40 2101 GTTGGGTTGG CGCCCGCGAA TCGCACTGAA AGACGGCATC GATGCAACGG TGTCGTGGTA
 2161 CCGCACAAAT GCCGATGCCG TGAGGAGGTA AAGCTGCGGG CCGGCCGATG TTATCCCTCC
 2221 GCGCGGACCG GTAGGGCGAC CTGCCATCGA GTGGTACGGC AGTCGCTCGG CCGCGGAGGC
 2281 GCATGGCCTA TGGGAGTATC CCATAGCCTG GCTTGGCTCG CCCCTACGCA TTATCAGTTG
 2341 ACCGCTTTTC CGCCAGCTCG CAGGCTCGCG GCAGCATCCC GTTCAGGTCT CCTCATGGTC
 45 2401 CCGTGTGGCA CGACCAACGA AGCTCGAACC GACTCGTTTC CCAATTTGCG ATGCTAATAT
 2461 CGCTCGATGG ATTTTTCGCG CAACGCCGGC TTGATGGCTC GTAACGTTAG CACCGAGATG
 2521 CTGCGCCACT TCGAACGAAA GCGCCTATTA GTAAACCAAT TCAAAGCATA CGGAGTCAAC
 2581 GTTGTATTG ATGTCCGTGC TAACTCCGGC CAGTTCGGTA GCGCTTTGCG TCGTGCAGGA
 2641 TTCAAGAGCC GTATCGTTTC CTTTGAACCT CTTTCGGGGC CATTTCGCGA ACTAACCGCG
 50 2701 GAGTCGGCAT CGGATCCACT ATGGGAGTGT CACCAGTATG CCTAGGCGA CGCCGATGAG

2761 ACGATTACCA TCAATGTGGC AGGCAATGCG GGGGCAAGTA GTTCCGTGCT GCCGATGCTT
 2821 AAAAGTCATC AAGATGCCTT TCCTCCCGCG AATTATATTG GCACCGAAGA CGTTGCAATA
 2881 CACCGCCTTG ATTCGGTTGC ATCAGAATTT CTGAACCCTA CCGATGTTAC TTTCCTGAAG
 2941 ATCGACGTAC AGGGTTTCGA GAAGCAGGTT ATCGCGGGCA GTAAGTCAAC GCTTAACGAA
 5 3001 AGCTGCGTCG GCATGCAACT CGAACTTTCT TTTATTCGT TGTACGAAGG TGACATGCTG
 3061 ATTCATGAAG CGCTTGAACT TGTCTATTCC CTAGGTTTCA GACTGACGGG TTTGTTGCCC
 3121 GGATTTACGG ATCCGCGCAA TGGTCGAATG CTTCAAGCTG ACGGCATTTT CTTCCGTGGG
 3181 GACGATTGAC ATAAATGCTT GCGTCGGCAC CCTGCCGGTA TCCAAACGGG CGATCTGGTG
 3241 AGCCGGCCTC CCGGGCACCT AATCGACTAT CTAAATTGAG GCGGCCGCGA CGTGCGGCAC
 10 3301 GAACAGGTGG CCGGCTGCTA GCGTTACACA CGTCATGACT GCGCCAGTGT TCTCGATAAT
 3361 TATCCCTACC TTCAATGCAG CCGTGACGCT GCAAGCCTGC CTCGGAAGCA TCGTCGGGCA
 3421 GACCTACCGG GAAGTGGAAG TGGTCCTTGT CGACGGCGGT TCGACCGATC GGACCCTCGA
 3481 CATCGCGAAC AGTTTCCGCC CGGAACCTCGG CTCGCGACTG GTGTTTCA GCGGGCCCGA
 3541 TGATGGCCCC TACGACGCCA TGAACCGCGG CGTCGGCGTA GCCACAGGCG AATGGGTACT
 15 3601 TTTTTTAGGC GCCGACGACA CCCTCTACGA ACCAACCACG TTGGCCCAGG TAGCCGCTTT
 3661 TCTCGGCGAC CATGCGGCAA GCCATCTTGT CTATGGCGAT GTTGTGATGC GTTCGACGAA
 3721 AAGCCGGCAT GCCGGACCTT TCGACCTCGA CCGCCTCCTA TTTGAGACGA ATTTGTGCCA
 3781 CCAATCGATC TTTTACCGCC GTGAGCTTTT CGACGGCATC GGCCCTTACA ACCTGCGCTA
 3841 CCGAGTCTGG GCGGACTGGG ACTTCAATAT TCGCTGCTTC TCCAACCGG CGCTGATTAC
 20 3901 CCGCTACATG GACGTCGTGA TTTCCGAATA CAACGACATG ACCGGCTTCA GCATGAGGCA
 3961 GGGGACTGAT AAAGAGTTCA GAAAACGGCT GCCAATGTAC TTCTGGGTG CAGGGTGGGA
 4021 GACTTGCAAG CGCATGCTGG CGTTTTTGAA AGACAAGGAG AATCGCCGTC TGGCCTTGCG
 4081 TACGCGGTTG ATAAGGGTTA AGGCCGCTC CAAAGAACGA AGCGCAGAAC CGTAGTCGCG
 4141 GATCCACATT GGACTTCTTT AACGCGTTTG CGTCCTGATC CACCTTTCAA CCCCCTTCCG
 25 4201 CGTGACGCGG CGCGCAGAGA GTGGTCGCAT ATCGCGTCAC TGTCTCTGTG CCAGTGCTTG
 4261 GAAAGCGTCG AGCACTCTGG TTCGCGTTCT TGACGTTTCG CCCCSCCCCT AGAGGTAGCG
 4321 TGTACGTGA CTGAAGCCAA TGAGTGCAAC TCGGCGTCGC GAAAGGTTTC AGTCGCGGTT
 4381 GAGCAAGACA CCGCAAGACT ACTGGAGTGC GTGCACAAGC GCCTCCAGCT CACGG

Seq. ID No.5

30 1 atgatcgctg tgatctggtc ggcgggtgccc acaggaaccg tcgacttgct gacgatcacc
 61 ttgtaccggt cgatgtatga cccaatgtcg tccgcaaccg agaagacgta cgtcaggctc
 121 gccgccccgc tttcacccat gggcgctcggg acggcgatga aaatgacgct cgcgtgctcg
 181 attccgcgct gccgggtcggg ggtgaagtca atcagcccg tctcagggc cctcgcaatc
 241 aactcccaac ccgggctcga aaatcgggac actgcctgcg aggagcaaat cgatcttggc
 35 301 ctgategata tcgacacaga cgacatcggt gccgctatcc gcgagacagg cgcccgtgac
 361 gaggcctaca tagcctga

Seq. ID No.6

40 1 M I A V I W S A V P T G T V D L S T I T L Y R S M Y D P M S
 31 S A T E K T Y V R S A A P L S P M G V G T A M K M T S A C S
 61 I P R C R S V V K S I S P F S R F L A I N S Q P G L E N R D
 91 T A C E E Q I D L G L I D I D T D D I V A A I R E T G A R D
 121 E A Y I A

Seq. ID No.7

1 gtgtcatctg ctccaaccgt gtcggtgata acgatttcgc tgaacgatcc cgagggattg
61 aaaagcaccg tggagagcgt tcgcgcgcag cgctatgggg ggcgaaatcga gcacatcgtc
121 atcgacggtg gatcggggcga cgccgtcgtg gagtatctgt ccggcgatcc tggctttgca
5 181 tactggcaat ctccagcccg caacgggaga tatgacgcga tgaatcaggg cattgcccat
241 tcgtcggggc accctgttctg gttctatgcac tccacggatc gtttctccga tccagatgca
301 gtcgcttccg tgggtggaggc gctctcgggg catggaccag tacgtgattt gtgggggttac
361 gggaaaaaca accctgtcgg actcgacggc aaaccacttt tccctcggcc gtacggctat
421 atgccgttta agatgcggaa atttctgctc ggcgcgacgg ttgcgcatca ggcgacattc
10 481 ttcggcgcgct cgctggtagc caagtggggc ggttacgac ttgatttttg actcgaggcg
541 gaccagctgt tcatctaccg tgcgcacta atacggcctc ccgtcacgat cgaccgcgtg
601 gtttgcgact tcgatgtcac gggacctggt tcaaccacgc ccatccgtga gcactatcgg
661 accctgcggc ggctctggga cctgcattgg gactaccgcg tgggtggggc cagagtgtcg
721 tgggcttact tgcgtgtgaa ggagtacttg attcggggcg acctggccgc attcaacgcg
15 781 gtaaaagtct tgcgagcgaa gttcgccaga gcttcggcga agcaaaattc atag

Seq. ID No.8

1 V S S A P T V S V I T I S L N D L E G L K S T V E S V R A Q
31 R Y G G R I E H I V I D G G S G D A V V E Y L S G D P G F A
61 Y W Q S Q P D N G R Y D A M N Q G I A H S S G D L L W F M H
20 91 S T D R F S D P D A V A S V V E A L S G H G P V R D L W G Y
121 G K N N L V G L D G K P L F P R P Y G Y M P F K M R K F L L
151 G A T V A H Q A T F F G A S L V A K L G G Y D L D F G L E A
181 D Q L F I Y R A A L I R P P V T I D R V V C D F D V T G P G
211 S T Q P I R E H Y R T L R R L W D L H G D Y P L G G R R V S
25 241 W A Y L R V K E Y L I R A D L A A F N A V K F L R A K F A R
271 A S R K Q N S

Seq. ID No.9

1 gtgaagcgag cgcttataac agggatcacg gggcaggatg gttcctacct cgccgagcta
61 ctactgagca agggatacga ggttcacggg ctcggttcgtc gagcttcgac gtttaacacg
121 tcgcggatcg atcacctcta cgttgaccca caccaaccgg gcgcgcgctt gttcttgca
30 181 tatgcagacc tcaactgacg caccgggttg gtgacctgc tcagcagrat cgaccgggat
241 gaggtctaca acctcgcagc gcagtcccat gtgcgcgtca gctttgacga gccagtgcac
301 accggagaca ccaccggcat gggatcgatc cgacttcttg aagcagtcgg cttttctcgg
361 gtggactgcc ggttctatca ggcttctctg tcggagatgt tcggcgcatc tccgccaccg
35 421 cagaacgaat cgacgcgctt ctatccccgt tcgccatacg gcgcggccaa ggtcttctcg
481 tactggacga ctcgcaacta tcgagaggcg tacggattat tcgcagtga tggcatcttg
541 ttcaaccatg agtcccccg gcgcggcgag acttctctga cccgaaagat cacgcgtgcc
601 gtggcgcgca tccgagctgg cgtccaatcg gaggtctata tgggcaacct cgatgcgac
661 cgcgactggg gctacgcgcc cgaatatgtc gaggggatgt ggaggatgtt gcaagcgctt
40 721 gaacctgatg actacgtcct ggcgacaggc cgtgggtaca ccgtacgtga gttcgtctaa
781 gctgcttttg accatgtcgg gctcgactgg caaaagcgcg tcaagtctga cgaccgctat
841 ttgcgtccca ccgaggtcga ttcgctagta ggagatgccg acaaggcgcc ccagtcactc
901 ggctggaaaag cttcgggtta tactgggtgaa ctgcgcgcga tcatgggtgga cgcggacatc
961 gccgcgtcgg agtgcgatgg cacaccatgg atcgacacgc cgatgttggc tgggtggggc
45 1021 agagtaagtt ga

Seq. ID No.10

1 VKRALITGITGQDGSYLAELLLSKGYEVHG
 31 LVRRASTFNTSRIDHLYVDPHQPGARLFLH
 61 YADLTDGTRLVTLSSIDPDEVYNLAQAQSH
 5 91 VRVSFDEPVHTGDTTGMGSI RLL EAVRLSR
 121 VDCRFYQASSSEMFGASPPPQNESTPFYPR
 151 SPYGA AKVFSYWTTRNYREAYGLFAVNGIL
 181 FNHESPRRGETFVTRK ITRAVARIRAGVQS
 211 EVYMG NLD AIRDWGYAPEYVEGMWRMLQAP
 10 241 EPDDYVLATGRGYTVREFAQA AFDHVG LLDW
 271 QKR VKFDDRYLRPT EVDSL VGDAADKAAQSL
 301 GWKASVHTGELARIMVDADIAALECDGTPW
 331 IDTPMLPGWGRVS

Seq. ID No.11

15 1 gtgaagcgag cgcttataac agggatcacg gggcaggatg gttcctacct cgccgagcta
 61 ctactgagca agggatacga ggttcacggg ctctctctgc gagcttcgac gtttaacacg
 121 tcgcgatcg atcacctcta cgttgaccca caccaaccgg gcgcgcgctt gttcttcgac
 181 tatgcagacc tcaactgacgg caccgggttg gtgacctgc tcagcagtat cgaccgggat
 241 gaggtctaca acctcgacgc gcagtcctcat gtgcgcgca gctttgacga gccagtgcat
 20 301 accggagaca ccaccggcat gggatcgatc cgacttcttg aagcagtcgc ctttctctcg
 361 gtggactgcc ggttctatca ggttctctcg tcggagatgt tcggcgcatc tccgccaccg
 421 cagaacgaat cgacgcgctt ctatccccgt tcgccatacg gcgcggccaa ggttctctcg
 481 tactggacga ctgcgaacta tcgagaggcg tacggattat tcgcagtga tggcatcttg
 541 ttcaaccatg agtccccccg gcgcggcgag acttctctga ccgaaaagat cagcgctgcc
 25 601 gtggcgcgca tccgagctgg cgtccaatcg gaggtctata tgggcaacct cgatgcgcat
 661 cgcgactggg gctacgcgcc cgaatatgtc gaggggatgt ggaggatgtt gcaagcgctt
 721 gaacctgatg actacgtcct ggcgacaggc cgtgggtaca ccgtacgtga gttcgtctaa
 781 gctgcttttg accacgtcgg gctcgaactg caaaagcag tcaagtttga cgaccgttat
 841 ttgcgcccc aagaggtcga ttcgctagta ggagatgccg acagggcggc ccagtcactc
 30 901 ggctggaaag cttcggttca tactggtgaa ctgcgcgca tcatggtgga cgcggacatc
 961 gccgcgtcgg agtgcgatgg cacaccatgg atcgacacgc cgatggtgcc tgggtggggg
 1021 ggagtaagtt ga

Seq. ID No.12

1 VKRALITGITGQDGSYLAELLLSKGYEVHG
 35 31 LVRRASTFNTSRIDHLYVDPHQPGARLFLH
 61 YADLTDGTRLVTLSSIDPDEVYNLAQAQSH
 91 VRVSFDEPVHTGDTTGMGSI RLL EAVRLSR
 121 VDCRFYQASSSEMFGASPPPQNESTPFYPR
 151 SPYGA AKVFSYWTTRNYREAYGLFAVNGIL
 40 181 FNHESPRRGETFVTRK ITRAVARIRAGVQS
 211 EVYMG NLD AIRDWGYAPEYVEGMWRMLQAP
 241 EPDDYVLATGRGYTVREFAQA AFDHVG LLDW
 271 QKH VKFDDRYLRPT EVDSL VGDAADRAAQSL
 301 GWKASVHTGELARIMVDADIAASECDGTPW
 45 331 IDTPMLPGWGGVS

Seq. ID No.13

1 gtgcgatggc acaccatgga tcgacacgcc gatgttgccg ggctggggca gagtaagtgt
61 acgactacac ctggggccctt ggaccgcgca acgcccgtgt atatcgccgg tcatcggggg
121 ctggtcggct cagcgctcgt acgtagattt gaggcgagg ggctcaccaa tctcattgtg
5 181 cgatcacgcg atgagattga tctgacggac cgagccgcaa cgcttgattt cgtgcttgag
241 acaagaccac aggtgatcat cgatgcggcc gcacgggtcg gcggcatcat ggccaataac
301 acctatcccg cggacttctt gtccgaaaac ctccgaatcc agaccaattt gctcgacgca
361 gctgtcgccg tgcgtgtgct gcggctcctt ttcttcgggt cgtcatgcat ctaccggaag
421 tacgtctcgc aacctatcca cgagagtgtt ttattgactg gccctttgga gccaccaaac
10 481 gacgcgtatg cgatcgccaa gatcgccggc atcctgcaag ttcaggcggc taggcgcca
541 tatgggctgg cgtggattct tgcgatgccg actaacctct acggaccgga cgacaacttc
601 tccccgtccg ggtcgcattt cttgccggcg ctcatccgtc gatattgagg agccaaagct
661 ggtggtgcag aagaggtgac gaattggggg accggtactc cgcggcgcg aactctgcat
721 gtcgacgac tggcgagcgc atgcctgttc cttttggaac atttcgatgg tccgaaccac
15 781 gtcaacgtgg gcaccggcgt cgatcacagc attagcgaga tcgcagacat ggtcgctaca
841 gcggtggtgt acatcgcgca aacacgttgg gatccaacta aacccgatgg aaccgcgcg
901 aaactattgg acgtctccgc gctacgcgag ttgggttggc gcccgcgaa cgcactgaaa
961 gacggcatcg atgcaacggt gtcgtggtac cgcacaaatg ccgatgccgt gaggaggtaa

Seq. ID No.14

1 VRWHTMDRHADVAWL GQSKLTTPG PLDR A
31 TPVYIAGHRGLVGSALVRRFEAE GFTNLIV
61 RSRDEIDLTDRAATFD FVSETRPQVI IDAA
91 ARVGGIMANN TYPADFLSENLR IQTNLLDA
121 AVAVRVPRLLFLGSSCIY PKYAPQPIHESA
15 151 LLTGPLEPTNDAYAI AKIAGILQVQAVRRQ
181 YGLAWISAMPTNLYGPGDNFS PSGSHLLPA
211 LIRRYEEAKAGGAEEVTNWGTGT PRRELLH
241 VDDLASACLFLEHFDGPNHVN VGTGV DHS
271 ISEIADMVATAVG YIGETRWDPTKPDGTPR
30 301 KLLDVSA LRELGW RPRIAL KDGIDATVSWY
331 RTNADAVRR

- 48 -

Seq. ID No.15

1 gtgcgatggc acaccatgga tcgacacgcc gatgttgccr ggctggggcg gagtaagttg
 61 acgactacac ctgggcctct ggaccgcgca acgcccgtgt atactgcccg tcatcggggg
 121 ctggctcggt cagcgctctg acgtagattt gaggcgagg ggctcaccaa tctcattgtg
 5 181 cgatcacgcy atgagattga tctgacggac cgagccgcaa cgtttgattt tgtgtctgag
 241 acaagaccac aggtgatcat cgatgcggcc gcacgggtcg gcggcatcat ggcaataaac
 301 acctatcccg cggactctct gtccgaaaac ctccgaatcc agaccaattt gctcgacgca
 361 gctgtcgccg tcgctgtgcc gcggctctct ttcctcggtt cgtcatgcat ctaccggaag
 421 tacgctccgc aacctatcca cgagagtgtt ttattgactg gccctttgga gccaccaaac
 10 481 gacgcgtatg cgatcgccaa gatcgccggt atcctgcaag ttcaggcggt taggcgccaa
 541 tatgggctgg cgtggatctc tgcgatgcg actaacctct acggaccggy cgacaacttc
 601 tccccgtccg ggtcgcatct cttgcggcg ctcattccgt gatatgagga agccaaagct
 661 ggtggtgcag aagaggtgac gaattggggg accggtaact cgcggcgcg acttctgcat
 721 gtcgacgac tcggcgagcg atgcctgttc cttttggaac atttcgatgg tccgaaccac
 15 781 gtcaacgtgg gcaccggcgt cgatcacagc attagcgaga tcgcagacat ggtcgctacg
 841 gcgggtgggt acatcggcga aacacgttgg gatccaacta aacccgatgg aaccgcgcg
 901 aaactattgg acgtctccgc gctacgcgag ttgggttggc gcccggaat cgcactgaaa
 961 gacggcatcg atgcaacggt gtcgtgtgtac cgcacaaatg ccgatgccgt gaggaggtaa

Seq. ID No.16

20 1 V R W H T M D R H A D V A W L G R S K L T T T P G P L D R A
 31 T P V Y I A G H R G L V G S A L V R R F E A E G F T N L I V
 61 R S R D E I D L T D R A A T F D F V S E T R P Q V I I D A A
 91 A R V G G I M A N N T Y P A D F L S E N L R I Q T N L L D A
 121 A V A V R V P R L L F L G S S C I Y P K Y A P Q P I H E S A
 25 151 L L T G P L E P T N D A Y A I A K I A G I L Q V Q A V R R Q
 181 Y G L A W I S A M P T N L Y G P G D N F S P S G S H L L P A
 211 L I R R Y E E A K A G G A E E V T N W G T G T P R R E L L H
 241 V D D L A S A C L F L L E H F D G P N H V N V G T G V D H S
 271 I S E I A D M V A T A V G Y I G E T R W D P T K P D G T P R
 30 301 K L L D V S A L R E L G W R P R I A L K D G I D A T V S W Y
 331 R T N A D A V R R

Seq. ID No.17

35 1 atggattttt tgcgcaacgc cggcttgatg gctcgtaacg ttagtaccga gatgctgcgc
 61 cacttcgaac gaaagcgcc attagtaaac caattcaaag catacggagt caacgttgtt
 121 attgatgtcg gtgctaactc cggccagttc ggtagcgctt tgcgtcgtgc aggatccaag
 181 agccgtatcg tttcccttga acctctttcg gggccatttg cgcaactaac gcgcaagtcg
 241 gcatcggtac cactatggga gtgtcaccag tatgccctag gcgacgccga tgagacgatt
 301 accatcaatg tggcaggcaa tgcgggggca agtagttccg tgcgtccgat gcttaaaagt
 40 361 catcaagatg cctttctctc cgcgaattat attggcaccg aagacgttgc aatacaccgc
 421 cttgatccgg ttgcatcaga atttctgaac cctaccgatg ttactttctt gaagatcgac
 481 gtacagggtt tcgagaagca gggtatcacg ggcagtaagt caacgcttaa cgaaagctgc
 541 gtcggcatgc aactcgaact ttcttttatt ccgttgtacg aagggtgacat gctgattcat
 601 gaagcgcttg aacttgctta ttccttaggt ttcagactga cgggtttgtt gcccgcttt
 45 661 acggatccgc gcaatggtcg aatgcttcaa gctgacggca tttctctccg tggggacgat
 721 tga

- 49 -

Seq. ID No.18

1 M D F L R N A G L M A R N V S T E M L R H F E R K R L L V N
 31 Q F K A Y G V N V V I D V G A N S G Q F G S A L R R A G F K
 61 S R I V S F E P L S G P F A Q L T R K S A S D P L W E C H Q
 5 91 Y A L G D A D E T I T I N V A G N A G A S S S V L P M L K S
 121 H Q D A F P P A N Y I G T E D V A I H R L D S V A S E F L N
 151 P T D V T F L K I D V Q G F E K Q V I T G S K S T L N E S C
 181 V G M Q L E L S F I P L Y E G D M L I H E A L E L V Y S L G
 211 F R L T G L L P G F T D P R N G R M L Q A D G I F F R G D D

10 Seq. ID No.19

1 atggattttt tgcgcaacgc cggcttgatg gctcgtaacg ttagcaccga gatgctgcgc
 61 cacttcgaac gaaagcgccct attagtaaac caattcaaag catacggagt caacgctgtt
 121 attgatgtcg gtgctaactc cggccagttc ggtagcgctt tgcgtcgtgc aggattccaag
 181 agcgcgtatcg tttcctttga acctctttcg gggccatttg cgcaactaac gcgcgagtcg
 15 241 gcatcggatc cactatggga gtgtcaccag tatgccctag gcgacgccga tgagacgatt
 301 accatcaatg tggcaggcaa tgcgggggca agtagttccg tgcgtccgat gcttaaaagt
 361 catcaagatg cctttcctcc cgcgaattat attggcaccg aagacgttgc aatacaccgc
 421 cttgattcgg ttgcatcaga atttctgaac cctaccgatg ttactttcct gaagatcgac
 481 gtacaggggt tgcagaagca gggttatcgcg ggcagtaagt caacgcttaa cgaaagctgc
 20 541 gtcggcatgc aactcgaact tctctttatt ccgttgtagc aaggtgacat gctgattcat
 601 gaagcgcttg aacttgctta ttccttaggt ttcagactga cgggtttgtt gcccggtatt
 661 acggatccgc gcaatggtcg aatgcttcaa gctgacggca tttctcttcg tggggacgat
 721 tga

Seq. ID No.20

25 1 M D F L R N A G L M A R N V S T E M L R H F E R K R L L V N
 31 Q F K A Y G V N V V I D V G A N S G Q F G S A L R R A G F K
 61 S R I V S F E P L S G P F A Q L T R E S A S D P L W E C H Q
 91 Y A L G D A D E T I T I N V A G N A G A S S S V L P M L K S
 121 H Q D A F P P A N Y I G T E D V A I H R L D S V A S E F L N
 30 151 P T D V T F L K I D V Q G F E K Q V I A G S K S T L N E S C
 181 V G M Q L E L S F I P L Y E G D M L I H E A L E L V Y S L G
 211 F R L T G L L P G F T D P R N G R M L Q A D G I F F R G D D

- 50 -

Seq. ID No.21

1 atgactgcgc cagtgttctc gataattatc cctaccttca atgcagcggc gacgctgcaa
 61 gcctgcctcg gaagcatcgt cgggcagacc taccgggaag tggaaagtgt ccttgtcgac
 121 ggcgggttcga ccgacgggac cctcgacatc gcgaacagtt tccgcccggg actcggctcg
 181 cgactgggtcg ttcacagcgg gcccgatgat ggcccttacg acgccatgaa ccgcggcgctc
 241 ggcgtggcca caggcgaatg ggtacttttt ttaggcgccg acgacacctt ctacgaacca
 301 accacgttgg cccaggtagc cgcttttctc ggcgaccatg cggcaagcca tcttgtctat
 361 ggcgatgttg tgatgcgttc gacgaaaagc cggcatgccg gacctttcga cctcgaccgc
 421 ctcttatttg agacgaattt gtgccaccaa tcgatctttt accgccgtga gcttttcgac
 481 ggcacgggcc cttacaacct gcgctaccga gtctgggcgg actgggacct caatattcgc
 541 tgcttctcca acccggcgct gattaccgcg tacatggacg tcgtgatttc cgaatacaac
 601 gacatgaccg gcttcagcat gaggcagggg actgataaag agttcagaaa acggctgcca
 661 atgtacttct ggggtgcagg gtgggagact tgcaggcgca tgctggcgct tttgaaagac
 721 aaggagaatc gccgtctggc cttgcgtacg cggttgataa gggttaaggc cgtctccaaa
 781 gaacgaagcg cagaaccgta g

Seq. ID No.22

1 M T A P V F S I I I P T F N A A V T L Q A C L G S I V G Q T
 31 Y R E V E V V L V D G G S T D R T L D I A N S F R P E L G S
 61 R L V V H S G P D D G P Y D A M N R G V G V A T G E W V L F
 91 L G A D D T L Y E P T T L A Q V A A F L G D H A A S H L V Y
 121 G D V V M R S T K S R H A G P F D L D R L L F E T N L C H Q
 151 S I F Y R R E L F D G I G P Y N L R Y R V W A D W D F N I R
 181 C F S N P A L I T R Y M D V V I S E Y N D M T G F S M R Q G
 211 T D K E F R K R L P M Y F W V A G W E T C R R M L A F L K D
 241 K E N R R L A L R T R L I R V K A V S K E R S A E P

Seq. ID No.23

1 atgactgcgc cagtgttctc gataattatc cctaccttca atgcagcggc gacgctgcaa
 61 gcctgcctcg gaagcatcgt cgggcagacc taccgggaag tggaaagtgt ccttgtcgac
 121 ggcgggttcga ccgacgggac cctcgacatc gcgaacagtt tccgcccggg actcggctcg
 181 cgactgggtcg ttcacagcgg gcccgatgat ggcccttacg acgccatgaa ccgcggcgctc
 241 ggcgtagcca caggcgaatg ggtacttttt ttaggcgccg acgacacctt ctacgaacca
 301 accacgttgg cccaggtagc cgcttttctc ggcgaccatg cggcaagcca tcttgtctat
 361 ggcgatgttg tgatgcgttc gacgaaaagc cggcatgccg gacctttcga cctcgaccgc
 421 ctcttatttg agacgaattt gtgccaccaa tcgatctttt accgccgtga gcttttcgac
 481 ggcacgggcc cttacaacct gcgctaccga gtctgggcgg actgggacct caatattcgc
 541 tgcttctcca acccggcgct gattaccgcg tacatggacg tcgtgatttc cgaatacaac
 601 gacatgaccg gcttcagcat gaggcagggg actgataaag agttcagaaa acggctgcca
 661 atgtacttct ggggtgcagg gtgggagact tgcaggcgca tgctggcgct tttgaaagac
 721 aaggagaatc gccgtctggc cttgcgtacg cggttgataa gggttaaggc cgtctccaaa
 781 gaacgaagcg cagaaccgta g

- 51 -

Seq. ID No.24

1 M T A P V F S I I I P T F N A A V T L Q A C L G S I V G Q T
 31 Y R E V E V V L V D G G S T D R T L D I A N S F R P E L G S
 61 R L V V H S G P D D G P Y D A M N R G V G V A T G E W V L F
 5 91 L G A D D T L Y E P T T L A Q V A A F L G D H A A S H L V Y
 121 G D V V M R S T K S R H A G P F D L D R L L F E T N L C H Q
 151 S I F Y R R E L F D G I G P Y N L R Y R V W A D W D F N I R
 181 C F S N P A L I T R Y M D V V I S E Y N D M T G F S M R Q G
 211 T D K E F R K R L P M Y F W V A G W E T C R R M L A F L K D
 10 241 K E N R R L A L R T R L I R V K A V S K E R S A E P

Seq. ID No.25

1 gtggccagca gaagccccca ctccgctgcg ggtgggtggc taattcttgg cggtcccttc
 61 cttgtggtcg gcgtggcgca tccggtagga ctccgccggag gtgacgacga tgctggcggtg
 121 gtgcagcagc cgatcgagga tgctggcggc ggtgggtgtg tcggggcagga atcgccccca
 15 181 ttgttcgaag ggccaatgcg aggcgatggc cagggagcgg cgctcgtagc cggcagccac
 241 gagccggaac aacagttgag tcccggtgtc gtcgagcggg gcgaagccga tctcgtccaa
 301 gatgaccaga tccgcgcgga gcagggtgtc gatgatcttg ccgacggtgt tgcgggccag
 361 gccgcggtag ~~aggacctgga~~ ccagggtggc ggcgggtgaag tagcggactt tgaatccggc
 421 gtggacggca gcgtgcccgc agccgatgag cagggtgactt ttgcccgtac cagggtgggccc
 20 481 aatgaccgcc aggttctgtt gtgcccgaat ccattccagg ctccgacaggt agtcgaacgt
 541 ggctgcggtg atcgacgacg cgggtgacgtc gaacccgtcg agggctcttg tgaccgggaa
 601 ggctgcggcc ttgagacggt tggcggtgtt ggaggcatcg cgggcagcga tctcggccctc
 661 aaccaacgtc cgcaggatct cctccggtgt ccagcgttgc gtcttggcga cttcgaacac
 721 ctcggcggtg ttgcggcgca ccgtggccag cttcaaccgc cgcagcggcg cgtcaaggtc
 25 781 agcagccagc ggtgcccgcg aggacgggtg caccggcttg gcagcgggtg tcatgaggcc
 841 gtcccgtcgg tgggtgtgat cttgtag

Seq. ID: No.26

1 V A S R S P H S A A G G W L I L G G S L L V V G V A H P V G
 31 L A G G D D D A G V V Q Q P I E D A G G G V L G Q E S P P
 61 L F E G P M R G D G Q G A A L V A G S H E P E Q Q L S P G V
 30 91 V E R G E A D L V Q D D Q I R A E Q G V D D L A D G V V G Q
 121 A A V E D L D Q V G G G E V A D F E S G V D G S V P A A D E
 151 Q V T F A R T R W A N D R Q V L L C P N P F Q A R Q V V E R
 181 G C G D R R S G D V E P V E G L G D R E G C G L E T V G G V
 35 211 G G I A G S D L G L N Q R P Q D L L R C P A L R L G D L Q H
 241 L G G V A A H R G Q L Q P P Q R R V K V S S Q R C R R G R C
 271 H R L G S G G H E A V P S V V L I L

- 52 -

Seq. ID No.27

1 atgggctgcc tcaaaggctgg tgcctgctgcc aatgttcttg tcccaacacc ggattatgtg
61 cgattcgcgt cccactatgg ctctctccg gacttcctgcc acggtgcgga tccgcaatcg
121 aagggcatcg tggagaacct ctgtggctac gctcaggacg accttgcggt gccgctgctg
5 181 accgaagctg cgttagccgg tgagcaggtc gacctacgtg ccttcaacgc ccaggcgcaa
241 ctatggctgc ccgaggtcaa tgccacggtc cactcggaga tctgcgccgt gcccacgat
301 cgcttggctg acgagcgcac cgtcttgagg gagctgcccc cgctgcggcc gacgatcggc
361 tcggggctcg tgcgcgtaa ggtcgacggc ctctcgtgca tccgttacgg ctacagctcgt
421 tactcggctg ctcagcggct cgtcggctgcc accgtggcgg tggcggctga tcatggcgcc
10 481 ctgatcctgt tggaaacctg gaccggctgt atcgtggcgg agcacgagct cgtcagccca
541 ggtgaggtgt ccatcctcga tgaacactac gacggaccca gaccgcacc ctgcgctggt
601 cctcgccccg aaaccaagc agagaaacga ttctgcgcac tgggaaccga agcgcagcag
661 ttccctcgtg gtgctgctgc gatcggcaac acccgactga aatccgaact cgacattctg
721 ctcgcccttg gcgcgcacca cggcgaaacg gctttgattg acgcgctgcg ccgggcgggt
15 781 gcgtttcggc ggttcgcgcg tgccgacgtg cgtctgaccc tggccgcggg cgccggcacc
841 ccacaacccc gccccgcggg cgacgcactc gtgctcgatc tgcccaccgt cgagaccgcg
901 tcgttgaggg cctacaagat caacaccacc gacgggacgg cctcatgacc accgctgcca
961 agccggctgg accgtcctcg gcggcaccgc tggctgctga ccttgacgcg gcgctgcggc
1021 ggttgaagct ggccacgggt cgccgcaacg ccgccgaggt gttgcaagtc gccaagacgc
20 1081 aacgctggac accggaggag atcctgcgga cgttggctga ggccgagatc gctgcccgcg
1141 atgcctccaa caccgccaac cgtctcaagg ccgacgccc cccggtcacc aagaccctcg
1201 acgggttcga cgtcaccgga tcgtcgatca ccgacgccac gttcgactac ctgtcgagcc
1261 tgggaatgat tcgggcacaa cagaacctgg cggctcattg cccacctggt acgggcaaaa
1321 gtcacctgct catcggtcgc gggcacgctg ccgtccacgc cggattcaaa gtccgctact
25 1381 tcaccgcccg cgacctgacg gaggtcctct accgcggccc ggccgacaac accgtcggca
1441 agatcatcga caccctgctc cgcgcggatc tggctacctt ggacgagatc ggcttcgccc
1501 cgtctgacga caccgggact caactgttgt tccggctcgt ggctgcccgc tacgagcgcc
1561 gctccctggc catcgctcgt cattggcccc tcgaacaatg ggggcgattc ctgcccagc
1621 acaccaccgc cgccagcatc ctcgatcggc tgctgcacca cgccagcatc gtcgtcacct
30 1681 ccggcgagtc ctaccggatg cgccacgccc accacaagaa gggagccgcc aagaattag

Seq. ID No.28

1 M G C L K G G V V A N V V V P T P D Y V R F A S H Y G F V P
31 D F C H G A D P Q S K G I V E N L C G Y A Q D D L A V P L L
61 T E A A L A G E Q V D L R A L N A Q A Q L W C A E V N A T V
91 H S E I C A V P N D R L V D E R T V L R E L P S L R P T I G
121 S G S V R R K V D G L S C I R Y G S A R Y S V P Q R L V G A
151 T V A V V V D H G A L I L L E P A T G V I V A E H E L V S P
181 G E V S I L D E H Y D G P R P A P S R G P R P K T Q A E K R
211 F C A L G T E A Q Q F L V G A A A I G N T R L K S E L D I L
241 L G L G A A H G E Q A L I D A L R R A V A F R R F R A A D V
271 R S I L A A G A G T P Q P R P A G D A L V L D L P T V E T R
301 S L E A Y K I N T T D G T A S

Seq. ID No.29

1 MTTAAKPVAPSSAAPLAADLDAALRRRLKLA
31 TVRRNAAEVLQVAKTQRWTP EEILRTLVEA
61 EIAARDASNTANRLKAAAFPVTKTLDGFDV
5 91 TGSSITAATFDYLSSEWIRAQONLAVIGP
121 PGTGKSHLLIGCGHAAVHAGFKVRYFTAAD
151 LIEVLYRGLADNTVGKIIDTLRLRADLVILD
181 EIGFAPLDDTGTQLLFRLVAAGYERRSLAI
211 ASHWPFEQWGRFLPEHTTAASILDRLHLHA
10 241 SIVVTSGESYRMRHADHKKGAAKN

Seq. ID No.30

1 gtgacgtctg ctccgaccgt ctcggtgata acgatctcgt tcaacgacct cgacgggttg
61 cagcgcacgg tgaaaagtgt gcgggcgcaa cgctaccggg gacgcacga gcacatcgta
121 atcgacggcg gcagcgggcg cgacgtggtg gcatacctgt ccgggtgtga accaggtctc
15 181 gcgtattggc agtccgagcc cgacggcggg cggtagcagc cgatgaacca gggcatcgcg
241 cagcatcggt gtgatctgtt gtggtctctg cactccgccc atcgtttttc cgggcccgcac
301 gtggtagccc aggccgtgga ggcgttatcc ggcaaggga cgggtgtccga attgtggggc
361 ttcgggatgg atcgtctcgt cgggctcgat cgggtgcgcg gcccgatacc tttcagcctg
421 cgcaaattcc tggccggcaa gcagggtgtt ccgcatcaag catcgttttt cggatcatcg
20 481 ctggtggcca agatcgggtg ctacgacctt gatttcggga tcgccgccga ccaggaattc
541 atattgcggg ccgcgctggt atgcgagccg gtcacgattc ggtgtgtgct gtgcgagttc
601 gacaccacgg gcgtcggctc gcaccgggaa ccaagcgcg tcttcgggtga tctgcgcgcg
661 atgggcgacc ttcacgcgcg ctaccggttc gggggaaggc gaatatcaca tgcctacctc
721 cgcggccggg agttctacgc ctacaacagt cgattctggg aaaacgtctt cagcggaatg
25 781 tcgaaatag

Seq. ID No.31

1 MTSAPT VSVITISFNDLDGLQRTVKSVRAQ
31 RYRGRIEHIVIDGGSGDDVVAYLSGCEPGF
61 AYWQSEPDGGRYDAMNQGIAHASGDLWLFL
30 91 HSADRFSGPDVVAQAVEALSGKGPVSELWG
121 FGMDRLVGLDRVRGPFPFSLRKFLAGKQVV
151 PHQASFFGSSLVAKIGGYDLDFGIAADQEF
181 ILRAALVCEPV TIRCVLC EFD TTGVGSHRE
211 PS AVFGDLRRMGDLHRRYPFGGRRI SHAYL
35 241 RGR E FYAYNSRFWENVFTRMSK

Seq. ID No.32

1 gtgaagcgag cgctcatcac cggaatcacc ggccaggacg gctcgtatct cgccgaactg
 61 ctgctggcca aggggtatga ggttcacggg ctcatccggc gcgcttcgac gttcaacacc
 121 tcgcggtatcg atcacctcta cgtcgacccg caccaaccgg gcgcggggtt gttctctcgc
 181 tatggtgacc tgatcgacgg aaccgggttg gtgacctgc tgagcaccat cgaacccgac
 241 gaggtgtaca acctggcggc gcagtcacac gtgcgggtga gcttcgacga acccgtgcac
 301 accggtgaca ccaccggcat gggatccatg cgactgctgg aagccgttcg gctctctcgg
 361 gtgcactgcc gcttctatca ggcgtcctcg tcggagatgt tcggcgcttc gccgccaccg
 421 cagaacgagc tgacgccgtt ctaccgcggc tcaccgtatg gcgcggccaa ggtctattcg
 481 tactggggca cccgcaatta tcgcgaagcg tacggattgt tcggcggtta cggcatcttcg
 541 ttcaatcacg aatcacccgg gcgcgggtgag acgttcgtga cccgaaagat caccagggcc
 601 gtggcacgca tcaaggccgg tatccagtcc gaggtctata tgggcaatct ggatgcggtc
 661 cgcgactggg ggtacgcggc cgaatacgtc gaaggcatgt ggcggatgct gcagaccgac
 721 gagcccgacg acttcgtttt ggcgaccggg cgcggtttca cctgcgtga gttcgcggcg
 781 gccgcgttcg agcatgccgg tttggactgg cagcagtacg tgaaattcga ccaacgctat
 841 ctgcggccca ccgaggtgga ttcgctgacg ggcgacgca ccaaggctgc cgaattgctg
 901 ggctggaggg cttcggtgca cactgacgag ttggctcgga tcatggctga cgcggacatg
 961 gcggcgctgg agtgcaagg caagccgtgg atcgacaagc cgatgatcgc cggccggaca
 1021 tga

20 Seq. ID No.33

1 MKRALITGITGQDGSYLAELLLAKGYEVHG
 31 LIRRASTFNTSRIDHLYVDPHQPGARLFLH
 61 YGDLIDGTRLVTL LSTIEPDEVYNLAAQSH
 91 VRVSFDEPVHTGDTTGMGSMRLLEAVRLSR
 121 VHC RFYQASSSEMFGASPPQNELTFYPR
 151 SPYGA AKVYSYWATRNYREAYGLFAVNGIL
 181 FNHESPRRGETFVTRKITRAVARIKAGIQS
 211 EVYMG NLD AVRDWGYAPEYVEGMWRMLQTD
 241 EPDDFVLATGRGFTVREFARAAFEHAGLDW
 271 QQYVKFDQRYLRPTEVDSLIGDATKAAELL
 301 GWRASVHTDELARIMVDADMALCEGKPW
 331 IDKPMIAGRT

Seq. ID No.34

1 atgaggctgg cccgtcgcgc tcggaacatc ttgcgtcgca acggcatcga ggtgctcgcg
 61 tactttgccg aactggactg ggaacgcaat ttcttgccc aactgcaatc gcatcgggtc
 121 agtgccgtgc tcgatgtcgg ggccaattcg ggcagtagc ccaggggtct gcgcggcgcg
 181 ggcttcgcgg gccgcatcgt ctcgctcgag ccgctgccc ggccctttgc cgtcttgacg
 241 cgcagcgcc ctacggaccc gttgtgggaa tgccggcgct gtgcgtggg cgatgtcgat
 301 ggaaccatct cgatcaacgt cgccggcaac gagggcgcca gcagttccgt cttgccgtc
 361 ttgaaacgac atcaggacgc cttccacca gccaaactacg tggcgcccca acgggtgccg
 421 atacatcgac tcgattccgt ggctgcagac gttctgccc ccaacgatat tgcttcttcg
 481 aagatcgacg ttcaaggatc cgagaagcag gtgatcgcg gtggcgattc aacgggtgcac
 541 gaccgatgcg tcggcatgca gctcgagctg tcttccagc cgttgtaga ggggtggcatg
 601 ctcatccgcg aggcgctcga tctcgtggat tcgtgggct ttacgtctc gggattgcaa
 661 cccggtttca ccgaccccg caacggctga atgctgcagg ccgatggcat cttcttcggg
 721 ggcagcgatt ga

Seq. ID No.35

1 MRLARRARNILRRNGIEVSRYFAELDWERN
 31 FLRQLQSHRVSAVL DVGANS GQYARGLRGA
 61 GFAGRIVSFEP L PGPFAVLQRSASTDPLWE
 5 91 CRRCALGDVDGTISINVAGNEGASSSVLPM
 121 LKRHQDAFP PANYVGAQRVPIHRLDSVAAD
 151 VLRPN DIAFLKIDVQGFEKQVIAGGDSTVH
 181 DRCVGMQLELSFQPLYEGGMLIREALDLVD
 211 SLGFTLSGLQPGFTDPRNGRMLQADGIFFR
 10 241 GSD

Seq. ID No.36

1 gtgaaatcgt tgaaactcgc tcgtttcctc gcgcgtagcg ccgccttcga gggttcgcgc
 61 cgctattctg agcgagacct gaagcaccag tttgtgaagc aactcaaact gcgtcggtta
 121 gatgtcgttt tcgatgtcgg cgccaactca ggacaatacg ccgcgcggcct ccgcgcgagca
 15 181 gcatataagg gccgcattgt ctcttcgaa ccgctatccg gaccgtttac gatcttgaa
 241 agcaaacgct caacggatcc actttgggat tgcgggcagc atgcgttggg cgattctgat
 301 ggaacggtta cgatcaatat cgcaggaaac gccggtcaga gcagtctcgt cttgcccattg
 361 ctgaaaagtc atcagaacgc ttttcccccg gcaaactatg tcggtaccca agaggcgctcc
 421 atacatcgac ttgattccgt ggccgcagaa tttctaggca tgaacggtgt cgctttcttc
 20 481 aaggctcgac ttcaaggctt tgaaaagcag gtgctcgccg ggggcaaact aaccatagat
 541 gaccattcgc tcggcatgca actcgaactg tccttcctgc cgttgtagca aggtggcatg
 601 ctcatctctg aagccctcga tctcgtgtat tccttgggct tcacgttgac gggattgctg
 661 ccttgcttca tggatgcaaa taatggtcga atgttgagg ccgacggcat ctttttcgcg
 721 gaggacgatt ga

25

Seq. ID No.37

1 MKSLKLARFIARSAAFEVSRRYSERDLKHQ
 31 FVKQLKSRRVDVVFDFTVGANS GQYAAGLR
 61 RAAYKGRIVSFEP LSGPFTILES KASTDPL
 91 WDCRQH ALGDS DGTVTIN IAGNAGQSSSVL
 121 PMLKSHQNAFP PANYVGTQEASI HRLDSVA
 151 PEF LGMNGVAFLKVDVQGFEKQVLAGGKST
 181 IDDHCVGMQLELSFLP LYE GGM LIPEALDL
 211 VYSLGFTLTGLLP CFIDANNGRMLQADGIF
 30 241 FREDD

Seq. ID No.38

1 atgggtgcaga cgaaacgata cgccggccttg accgcagcta acacaaagaa agtcgccatg
61 gccgcaccaa tgttttcgat catcatcccc accttgaacg tggctgcggt attgcctgcc
121 tgcctcgaca gcacgcgccg tcagacctgc ggtgacttcg agctggctact ggtcgacggc
5 181 ggctcgacgg acgaaaccct cgacatcgcc aacattttcg cccccaacct cggcgagcgg
241 ttgatcattc atcgcgacac cgaccagggc gtctacgacg ccatgaaccg cggcgtggac
301 ctggccaccg gaacgtgggt gctctttctg ggcgaggacg acagcctgta cgaggctgac
361 accctggcgc ggggtggcgc cttcattggc gaacacgagc ccagcgatct ggtatatggc
421 gacgtgatca tgcgctcaac caatttcgc tgggtggcg ccttcgacct cgaccgtctg
10 481 ttgttcaagc gcaacatctg ccacaggcg atcttctacc gccgcggact cttcggcacc
541 atcggctccct acaacctccg ctaccgggtc ctggccgact gggacttcaa tattcgctgc
601 ttttccaacc cagcgtcgt caccgctac atgcacgtgg tcgttgcaag ctacaacgaa
661 ttcggcgggc tcagcaatac gatcgtcgac aaggagtttt tgaagcggct gccgatgtcc
721 acgagactcg gcataaggct ggtcatagtt ctggtgcgca ggtggccaaa ggtgatcagc
15 781 agggccatgg taatgcgcac cgtcatttct tggcggcgcc gacgttag

Seq. ID No.39

1 M V Q T K R Y A G L T A A N T K K V A M A A P M F S I I I P
31 T L N V A A V L P A C L D S I A R Q T C G D F E L V L V D G
61 G S T D E T L D I A N I F A P N L G E R L I I H R D T D Q G
91 V Y D A M N R G V D L A T G T W L L F L G A D D S L Y E A D
121 T L A R V A A F I G E H E P S D L V Y G D V I M R S T N F R
151 W G G A F D L D R L L F K R N I C H Q A I F Y R R G L F G T
181 I G P Y N L R Y R V L A D W D F N I R C F S N P A L V T R Y
211 M H V V V A S Y N E F G G L S N T I V D K E F L K R L P M S
241 T R L G I R L V I V L V R R W P K V I S R A M V M R T V I S
25 271 W R R R R

Seq 40:

GATGCCCTGAGGAGGTAAAGCTGC

Seq 41:

30 GATACGGCTCTTGAATCCTGCACG